

Two-Way and Monolingual English Immersion in Preschool Education:
An Experimental Comparison

W. Steven Barnett, Ph.D.

Donald J. Yarosz, Ed.D.

Jessica Thomas, M.A.

Dulce Blanco, M.A.

National Institute for Early Education Research
Rutgers—The State University of New Jersey

Corresponding Author:

Steve Barnett

(932) 732-4350

NIEER

120 Albany St., Suite 500

New Brunswick, NJ 08901

sbarnett@nieer.org

Two-Way and Monolingual English Immersion in Preschool Education:
An Experimental Comparison

Abstract

An experimental study was conducted comparing the effects of a dual language or two-way immersion (TWI) and monolingual English immersion (EI) preschool program on children's learning. Three and four-year old children were randomly assigned by lottery to either a newly established TWI Spanish/English (n=85) program or a monolingual English (n=65) program in the same district. Children in the study were from both Spanish and English home language backgrounds. All classrooms in the study used the High/Scope curriculum. The TWI program alternated between English and Spanish weekly by rotating children between two classrooms (and teachers) each week. Programs were compared on measures of children's growth in language, emergent literacy, and mathematics. Among the native Spanish speakers, the TWI program produced large and significantly significant gains in Spanish vocabulary compared to the EI program. Results were somewhat more mixed on measures of literacy skills. Children in the TWI program made greater phonological awareness gains in both English (all children) and Spanish (comparison included native Spanish speakers only), while in the EI program native Spanish speakers made greater gains on letter recognition in Spanish. The TWI program did not require additional expenditures. On balance, preschool TWI appears to be a promising and a cost-effective approach to improving the education of both English language learners and native English speaking children.

Two-Way and Monolingual Immersion in Preschool Education:
An Experimental Comparison

Introduction

The number of children in the United States who are English language learners (ELL) is growing rapidly, and the vast majority are from homes where Spanish is the primary language (Tabors, Paez & Lopez, 2003). This trend is even stronger at the preschool level. Hispanic children account for more than 20% of all children under five (Collins & Ribeiro, 2004). Moreover, Hispanic children are more highly represented in public preschool programs because they are more likely to be income eligible for such programs. About 30% of the Head Start population is Hispanic (Collins & Ribeiro, 2004). In the public schools, about 25% of preschool children are Hispanic (Smith, Kleiner, Parsad & Farris, 2003). Head Start data indicate that for $\frac{3}{4}$ of these children Spanish is the dominant language (Collins & Ribeiro, 2004). These children present challenges and opportunities for public preschool programs (Barnett, Hustedt, Robin, & Schulman, 2004). A number of studies have found preschool education to improve the school readiness of Hispanic children who experience high rates of school failure and high school dropout (Gormley, Gayer, Phillips, & Dawson, 2004; Johnson & Walker, 1991; Puma, Bell, Cook, Heid, Lopez, Zill, et al., 2005).

A key issue in the education of ELL children is language of instruction. School systems historically have encouraged English language abilities to the neglect of a child's home language (Hakuta, 1986). However, a variety of bilingual approaches have been developed as a means to increase the achievement of ELL children. Research tends to find that bilingual and first language education are at least as effective as English

immersion, but these approaches remain controversial (Willig, 1985; Greene, 1997; Rossell & Baker, 1996; Slavin & Cheung, 2005). Some have suggested that children may lose their first language when a “high-prestige” second language is introduced, arguing for emphasis on the first language in preschool education (Wong Fillmore, 1991). Yet, studies have found that bilingual preschool programs promoted development in both languages rather than impeding growth in the first language (Winsler, Diaz, Espinosa, & Rodriguez, 1999). The outcomes appear to depend on how well development in each language is supported (Tabors & Snow, 2001).

Two-way bilingual (TWI) education is a promising and increasingly common approach to addressing the needs of English Language Learners (Short 1993; Howard, Sugarman, & Christian, 2003; Lindohom- Leary, 2005). TWI (also called dual language) programs provide ELL and native English speakers with an education in two languages. Such programs build on research on effective practices in the education of ELL students, including evidence that first language proficiency strengthens second language acquisition and that bilingualism has cognitive benefits (Cummins, 1981; Reynolds, 1991; August & Hakuta, 1997). Although much of this research has been conducted with older children, there is similar evidence from the preschool years (Cambell & Sais, 1995; Campos, 1991; Diaz, Padilla, & Weathersby, 1991). TWI programs also seek to redress the problem of America’s traditionally weak foreign language programs (Howard, Sugarman, & Christian, 2003).

Nevertheless, there is an urgent need for additional rigorous research to fully specify the most effective approaches to preschool education for ELL children, particularly the large number speaking Spanish (Bialystok and Herman, 1999; Garcia,

2000; Dickinson & Tabors, 2001; Tabors & Snow, 2001). Past studies with preschool children have suggested that bilingual, and TWI approaches, in particular, may have important advantages (e.g., Rodriguez, Diaz, Duran, & Espinosa, 1995; Winsler, et al., 1999; Stipek, Ryan, & Alarcon, 2001; Tabors, Paez, & Lopez, 2003). To our knowledge, the study presented here is the first to use random assignment of both ELL and other children to either a TWI or an English immersion (EI) preschool program to compare the educational effectiveness of these two approaches.

Method

The study was conducted in a Northeastern city with a population of just over 120,000 in which 50% of the population is Latino (U.S. Census Bureau, 2000). Median household income was \$35,175 in 1999, and 21% of the children under age 5 were in poverty. In 2004-05, 76% of the public school children qualified for free or reduced price lunch. One school in the city was selected to implement a new TWI preschool program. Otherwise children attended programs that were primarily in English (EI), though some support was provided in these programs for home language development in Spanish and other languages. A lottery for admission to the TWI program (because city wide demand for the program greatly exceeded the supply of spaces) created the opportunity for randomization.

Sample and Program Assignment

In the fall of 2002, nearly 1000 3- and 4-year olds applied to the TWI program. All applicants were entered into a lottery to select 150 3-year olds and 150 4-year olds for attendance. Each applicant had an equal chance (within the two age groups) to attend. We randomly selected 50 of the lottery “winners” at each age for the TWI program

group. Of these, 85 agreed to participate in the study (i.e., permit their children to be tested, etc.). From the list of lottery entrants not assigned to the TWI program and locatable, we randomly selected 127 for invitation to the study. Of these, 62 families agreed to participate. The lower percentage of participation by control (EI) families may be due to disappointment at not winning admission to the TWI program. The number of children tested at posttest as well as pretest was somewhat smaller (79 TWI and 52 EI) primarily because children moved from the programs (one EI child declined testing).

Descriptive statistics for the initial sample are displayed in Table 1. Despite the limitations of the sampling, the TWI and EI groups did not significantly differ in their demographic characteristics (independent samples t-tests) or on any of the pretests (as determined by ANOVA). Nevertheless, analyses of the posttest scores presented later control for age, language, and pretest to increase statistical power and because modest (but not significant) differences between TWI and EI groups on these variables might tilt the analyses.

Program Description

In the TWI program, each classroom of students rotates weekly between two teaching teams. One team consists of a teacher and an assistant who teach in English. The other team consists of a teacher and an assistant who teach in Spanish. Thus, each week half of the students in the school are taught in English and half are taught in Spanish. Members of both teaching teams collaborate to plan curricula and share assessment data for the children they jointly teach. Students who speak primarily English are integrated with students whose first language is not English. In our sample nearly 60% of children spoke primarily Spanish at home (SHL), 35% spoke primarily English

(EHL), and 7% had another primary language. Study children were enrolled in 20 (10 English and 10 Spanish) TWI classrooms and 16 EI classrooms.

All classrooms in the study employed a licensed teacher and an assistant with a Child Development Associate (CDA) credential to work with a maximum of fifteen children per classroom. All classrooms used the High Scope Curriculum (Hohmann and Weikart, 1995), which has been adopted citywide by the public schools. For comparison purposes, data were collected on teachers and on classroom quality for all 20 TWI classrooms and all 16 EI classrooms. Teachers did not significantly differ in experience, field of certification, or whether they took the traditional or alternate route to teacher certification. However, TWI teachers were more likely to have a Masters degree (30% v. 6%) and to be bilingual (50% v. 25%).

Child Assessment

Children were tested in the fall and spring of the 2003-04 school year. Fall testing took place over 5 weeks and spring testing spanned 4 weeks. The TWI group was administered each measure in both English and Spanish. In the EI group, SHL children were assessed in English and Spanish; the others were tested in English only. Children tested in both languages were tested first in their home language. Primary language was ascertained from the classroom teacher who made judgments based on language proficiency tests administered by the schools at the beginning of the year, supplemented by parental report and their own experiences with the children. Assessors spoke to the child only in the language of each assessment to avoid code switching during testing sessions. Assessments were conducted one-on-one in the child's school, and assessments

were scheduled to avoid meal, nap and outdoor play times. Testing sessions lasted 20-30 minutes.

For assessment and data analysis, children who had a primary language other than English were combined with the native English speakers in the EHL group. Ideally, these children would have constituted a separate group for assessment and analysis. However, given the small number this was not feasible, and the alternative would have been to delete these children from the study resulting in less statistical power. These children were found to have sufficient proficiency in English to participate in the English language assessments and did not significantly differ from the other EHL children on the pretest measures of language and literacy.

There were two teams of assessors, one for each language. They included research associates with years of experience testing children and less experienced recent college graduates and advanced teacher education students. Assessors were trained on each assessment and then shadow scored in practice assessments until they reached 100% agreement with the trainer, who also was the site coordinator. The site coordinator was responsible for assuring adequate reliability throughout the study.

Individualized assessments were selected to measure the contributions of the preschool programs to children's learning, with emphasis on language and emergent literacy skills. Criteria for instrument selection included: (1) availability of equivalent tasks in both Spanish and English, (2) reliability and validity, particularly pre-literacy skills that are good predictors of later reading ability; and (3) appropriateness for children ages 3 to 5. Each assessment instrument is discussed below.

The Peabody Picture Vocabulary Test (PPVT – III) and Test de Vocabulario en Imagenes Peabody (TVIP)

The PPVT – III (Dunn & Dunn, 1997) is a 204-item test of receptive vocabulary in standard English. The test is administered by having children point to one of four pictures shown when given a word to identify. The PPVT-III is often used as a quick indicator of general cognitive ability, and it correlates reasonably well with other measures of linguistic and cognitive development related to school success. The PPVT-III has a mean standard score of 100 and a standard deviation of 15. The PPVT-III was nationally standardized on a stratified normative sample of 2000 children and adolescents and has an internal consistency reported as Spearman-Brown split half reliability coefficients ranging from .92 to .98. Test-retest reliability for a one-month interval in four different age groups ranged from .91 to .93.

The TVIP (Dunn, Lugo, Padilla, & Dunn, 1986) uses 125 translated items from the PPVT to assess receptive vocabulary acquisition of Spanish-speaking and bilingual students. Items were selected through item analysis for their universality and appropriateness to Spanish-speaking communities, according to publishers. The test is appropriate for measuring growth in Spanish vocabulary for bilingual students and for those who do not come from Spanish language homes, according to its authors. The TVIP has a mean standard score of 100 and a standard deviation of 15. The TVIP was standardized on two monolingual, Spanish-speaking populations (1,219 children from public schools in Mexico and 1,488 children from the San Juan metropolitan area of Puerto Rico). Internal consistency reliability for this measure is reported to be .91 to .94

in the relevant age range. Correlations between the TVIP and Spanish language tests of general ability range from .25 to .56.

The Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R) and the Bateria Psico-Educativa Revisada de Woodcock-Muñoz (WM-R)

The WJ-R and WM-R (Woodcock & Johnson, 1989; Woodcock & Munoz-Sandoval, 1996) are comprehensive sets of individually administered tests of cognitive abilities and achievement. We administered the Picture Vocabulary and Applied Problems subtests from these batteries in both English and Spanish. The 58-item Picture Vocabulary (Vocabulario con Dibujos) subtest asks children to select pictures to match words and to say a word when shown a picture. It is primarily an expressive vocabulary task that measures the ability to name familiar and unfamiliar pictured objects. The 60-item Applied Problems (Problemas Aplicados) subtest measures math skills.

The WJ-R and WM-R tests were calibrated and equated to U.S. norms through Rasch modeling and are particularly well-suited to the needs of assessment with bilingual populations (Woodcock & Munoz-Sandoval, 1996). The tests' standard scores have a mean of 100 and a standard deviation of 15. The English Form of the subtests was normed on a stratified random sample of 6,359 English-speaking subjects in the United States. The Spanish Form was normed on 3,911 primarily monolingual Spanish-speaking subjects from samples obtained both inside and outside the United States. Internal consistency reliabilities range from the high .70s to low .90s on both subtests for preschool-aged children. Correlations of the WJ-R and WM-R with other tests of cognitive ability and achievement are reported to range from .60 to .70.

Measures of Emergent Literacy Skills

We administered two phonological awareness (phoneme deletion and rhyme recognition) subtests in both English and Spanish. These criterion-referenced tests are early versions of tests developed specifically to fill the need for equivalent tests in English and Spanish that could be used to study the literacy development of bilingual young children (Lopez & Greenfield, 2004; Lopez, 2005). The tests measure early skills that are highly predictive of later reading abilities and are important milestones in the literacy development of 3- and 4-year-old children (Snow, Burns, & Griffin, 1998). Phonological awareness and letter recognition are believed to play key roles in children becoming successful readers (Scarborough, 1998; Whitehurst & Lonigan, 1998).

The *Phoneme Deletion* test consists of three practice items followed by eight test items. A puppet is introduced and the child is shown a target picture and given its name (e.g., this is a rock). The child is asked to repeat the name of the picture and then the puppet repeats the name (sometimes correctly, sometimes deleting beginning or ending sounds). The child has to judge if the puppet is correct or not and then correct the puppet, if necessary. The child must pass at least one of the three practice items for the subtest to be administered. This test assesses a child's ability to discern correct versus incorrect pronunciation of basic vocabulary in both English and Spanish. The Spanish and English language scores on this test were moderately correlated, .49 at pretest and .32 at posttest. The Spanish language test correlated with the TVIP .38 at pretest and .47 at posttest. The English language test correlated with the PPVT .55 at pretest and .48 at posttest.

The *Rhyme Recognition* subtest consists of three practice items followed by eight test items. Children are shown three pictures on a page and are asked to point to the two that rhyme after repeating all of them first with the examiner. The child needs to pass at least one of the practice items in order for the subtest to be administered. The rhyme recognition tests proved too difficult for most children at pretest and remained difficult for many children at posttest when 51% could not complete the task in Spanish and 65% could not complete the task in English.

An *Alphabet Recognition* test was developed at NIEER to assess letter recognition in English and Spanish. This test allows children to identify all of the letters that they might know; other letter recognition tests commonly present only a small subset of the alphabet to children. A child is presented two letters of the alphabet on each page and is asked to name one of them at a time. The answer is considered correct if the child gives a correct letter name response, a correct letter sound response or a word starting with the letter presented. The score is calculated as the total number of correct responses. The Spanish and English language test scores correlated .78 at pretest and .93 at posttest.

Measurement of Classroom Environment and Activities

To identify similarities and differences between the two types of preschool programs we conducted classroom observations using three rating scales. Observers were trained on the classroom measures and then “shadow scored” on practice observations until they reached an eighty-five percent inter-rater reliability on each instrument with the trainer. All observations were conducted during February and March. All TWI classrooms (N=20) and EI classrooms (N=16) were observed.

Early Childhood Environmental Rating Scale- Revised (ECERS-R)

The ECERS-R (Harms, Clifford and Cryer, 1998) provides a global measure of preschool classroom quality with 43 items that cover a broad range of quality considerations from safety to teacher-child interaction to parent involvement. The ECERS-R scale uses a scoring system where 1 is considered inadequate, 3 minimal, 5 good, and 7 excellent. The ECERS- R has been widely used to measure quality in research studies within the United States and abroad. Scores have been found to correlate moderately with more specific measures of teaching quality and to predict children's learning and development (Peisner-Feinberg & Burchinal, 1997; Sylva & Siraj-Blatchford, 2001). The total scale internal consistency for the ECERS-R is .92, and subscale internal consistencies range from .71 to .88.

Supports for Early Literacy Assessment (SELA)

The SELA (Smith, Davidson, Weisenfeld and Katsaros 2001) was used to measure the quality of the preschool literacy environment and instruction in greater detail. Scores on the SELA range from 1-5, with 1 considered very low quality, 3 fair quality, and 5 ideal quality. The SELA was developed based on current research and professional opinion regarding best practices with regard to literacy. In another study, with 310 classrooms, the correlation between the SELA and ECERS-R total scores was found to be .75 (Lamy, 2004).

Supports for English Language Learners Classroom Assessment (SELLCA).

The SELLCA was developed to assess the extent to which the preschool teacher and classroom environment support the L1 development of ELL children. Scores on the SELLCA range from 1-5, with 1 considered very low, 3 fair, and 5 ideal. We used three

items from the SELLCA to measure teacher practices in all classrooms as an indication of the extent to which actual treatment corresponded to intent (TWI v. EI) and to assess differences between the TWI and EI programs. These three items related to: teachers' use of children's dominant language, degree to which children's cultural backgrounds are incorporated in the classroom, and use of effective strategies to support English language development. The SELLCA is a newly developed instrument (NIEER, 2005).

Results

Classroom Quality

As reported in Table 2, the average score was “good” on the ECERS-R and “fair” on the SELA. The ECERS-R scores are better than has been found in larger studies of preschool programs across the United States (NCEDL, 2005). T-tests revealed no significant differences in mean ECERS-R or SELA scores between TWI and EI group classrooms. As expected, the SELLCA found that the Spanish language TWI classrooms used Spanish much more frequently than did the EI classrooms. Also, the Spanish language TWI classrooms scored much better than the EI classrooms on incorporating the cultural background into classroom life. TWI and EI classrooms did not significantly differ on use of effective strategies to help children understand English. The SELLCA revealed that the Spanish language TWI classrooms used Spanish most (but not all) of the time while the EI classrooms used Spanish much less frequently, but did not entirely exclude support for children in Spanish.

Children's Learning

Regression analyses were employed to estimate the effects on all English language measures of independent variables for treatment (EI v. TWI), language, the

interaction between treatment and language, child age and pretest scores. As no statistically significant interaction effects were found, results are presented in Table 3 for simplified regressions. These analyses use raw scores. Standard scores are not available for all measures, but analyses of standard scores where available produced essentially the same results. At conventional levels of significance ($p < .05$) no program differences were found on English language and literacy development. The only indication of a possible program effect on English language measures was a small advantage for the TWI program group on English Phoneme Deletion, significant at the .06 alpha level. .

Regression analyses were performed on all Spanish language measures using only the sample children who were Spanish language dominant. These analyses estimated the effects of treatment (TWI v. EI), child age, and pretest scores. Results for raw score analyses are displayed in Table 4. (Again, results with standard scores were the same.) The Spanish language children in the TWI program group gained significantly and substantially more on the TVIP (7.5 standard score points, $p < .001$) controlling for age and Fall TVIP score compared to their peers in the EI group. Over the school year, TVIP standard scores (i.e., age adjusted scores) improved for the TWI group, but declined for the EI group. Spanish language children in the TWI program group also performed modestly, but significantly better, on the Spanish language rhyme test. One comparison favored the EI program group. Spanish language children in the EI program group scored modestly, but significantly higher than their TWI peers, on Spanish letter recognition, though the difference was modest (11 v. 8 letters).

One issue in any study with multiple outcome measures is that multiple statistical tests are conducted and this increases the risk of a Type I error. One approach to taking

this into account is to apply the Bonferroni correction to adjust the alpha level for statistical significance. The Bonferroni procedure is to divide the alpha level by the number of statistical tests conducted, in this case $.05/12$, or $.004$. The effect on the TVIP would be considered statistically significant. The significance of effects for the other outcomes is less robust, since none are significant if alpha is set at $.004$. However, the Bonferroni correction is overly conservative where the outcome measures are correlated, as is the case here.

Finally, the relative effects of TWI on Spanish language development for English dominant children could not be estimated in the same way because EI children with English (or another non-Spanish language) as a primary language were not tested in Spanish. However, it was possible to look at language development for English dominant children in the TWI classrooms over the year. Matched pairs t-tests were used to determine whether TWI children with English as their primary language made gains in Spanish from Fall to Spring. Statistically significant gains were found in Spanish receptive vocabulary, Spanish phoneme deletion, and Spanish letter recognition.

Discussion

This study used random assignment to compare the effects of TWI (Spanish and English) and EI (English) preschool programs. Most children served were Hispanic and spoke Spanish as the primary language at the home. The TWI program provided better support for Spanish language development without sacrificing gains in English language development. In addition, English dominant children in the TWI preschool program also made gains in Spanish language and literacy skills without hindering the development of their English language abilities (though this last finding is not bolstered by comparison to

a control group). While our outcomes are strictly short term, the research on the benefits of bilingualism reviewed earlier in the paper would suggest that longer term effects could be even more substantial.

We did not conduct a detailed cost analysis or benefit-cost analysis of the two programs because the costs of these programs are essentially the same. The TWI program did not require additional resources. It accomplishes its goals by hiring native Spanish speakers as lead teachers for half of the preschool classrooms rather than by adding additional staff to existing classrooms. With equal costs, the TWI program is a clear winner in economic terms given its superior effectiveness. The practical importance of this conclusion can not be overestimated. The TWI program demonstrates the ability to improve the education of ELL children and their native English speaking peers without increasing costs.

By far the largest difference observed between programs was that Spanish dominant children in the TWI program made much greater gains in Spanish vocabulary (receptive language). While the Spanish dominant children in the EI program lost ground during the year relative to age norms, their peers in the TWI program gained ground relative to age norms. The decline in TVIP standard scores for Spanish dominant children in the EI program is further evidence that immersion in an English-only preschool setting is accompanied by, and might even contribute to, Spanish language loss. The TVIP standard score gains for Spanish dominant children in the TWI program suggest that TWI preschool programs can ameliorate this problem with no loss of effectiveness in promoting English language and literacy development.

Two of three smaller effects identified favored the TWI program on phonological awareness. Spanish dominant children had greater gains on the Spanish language rhyme test in the TWI program than in the EI program. All TWI children may have out gained their EI peers ($p=.06$) on phoneme deletion in English. This suggests an overall advantage for the TWI program on skills that are important for future reading success in both languages (Bialystok, Majumder, & Martin, 2003; Tabors, Paez & Lopez, 2003).

The third finding favored the EI program, as Spanish dominant children in the EI program had greater gains in Spanish letter identification, and was unexpected. Perhaps the EI program spent more time on letter recognition in English, Spanish, or both languages. Letter identification skill may transfer readily from English to Spanish at this age, and sound-to-print mappings are more consistent in Spanish. The SELCA scores indicate that the EI program offered some supports for Spanish language development. However, scores specifically on the SELCA letter knowledge item do not support the hypothesis that there was more emphasis on letter identification skills in the EI program.

The emergent literacy skills tests used here were newly developed, and data on their performance is limited. It is possible that we found effects in one language but not another on some of these measures because the measures are not precisely equivalent in each language. Alternatively, preschool children may find it easier to demonstrate gains in some of these skills in one language than in another. For example, rhyme may be more salient in 1-syllable English words than in two-syllable Spanish words, even for native Spanish speakers. Program impacts might be more consistent across languages with older children with more mature cognitive abilities (Riccio, Amado, Jimenez, Hasbrouck, Imhoff, & Denton, 2001; English, Leafstedt, Gerber & Villaruz, 2001).

The three modest effects on emergent literacy skills just discussed should be interpreted with caution, given the multiple comparisons. Using the Bonferroni correction, these might be dismissed as too likely to have occurred by chance. In our view that would be an overly cautious approach. Taken together they add to the evidence that the TWI approach was more effective, even if not superior in every respect. These findings should at the very least stimulate future research on the effects of TWI and other approaches to preschool education on emergent literacy skills in English and Spanish (and other languages). In addition, future research could benefit from detailed qualitative descriptions of specific within-program literacy supports to obtain more information about the details of program implementation including, for example, the degree to which each program emphasized letter knowledge over other features of literacy support.

Future research also would benefit from further development of comparable measures in English and Spanish (and other languages) for the language and literacy development of young children. This study sought to measure growth in both languages, and we have encountered several limitations that others also have noted (Kester & Pena, 2002; Pearson, Fernandez, and Oller, 1993; Restrapo & Silverman, 2001; Tabors, Paez & Lopez, 2003). Better measures for young children will increase what can be learned from studies in this vital area of native and second language and literacy development.

The study reported here has limitations. It is modest in size and duration and limited to one particular location and population. Results might differ for children with other languages or even for Spanish dominant children with different backgrounds. The number of English dominant children was quite small, and the number of children with a third primary language was too small for separate analyses. In addition, the

treatments were not pure. By design, the TWI program was 100% immersion in each language on alternate weeks and the EI program was 100% English language immersion. In practice, the differences between programs were less extreme. The divergence from pure immersion models may have been adaptive or may have detracted from the effectiveness of one or both of the models. Future research can address these and other limitations. No single study should be expected to provide an adequate basis for policy and practice.

That said, it is clear that the TWI preschool program supported stronger Spanish language gains at no expense to English language development in both native English and native Spanish speakers. The study's methodology provides confidence that the findings accurately represent program effects. TWI appears to be a promising, cost-effective approach to improving the education of ELL and native English speaking children. In view of these findings, high priority should be assigned to further development and evaluation of TWI approaches to preschool education. More randomized trials are required to investigate how effectiveness may vary with specific program characteristics, populations served, community and school contexts, and other conditions that might affect success. TWI has the potential to significantly address two serious problems--inadequate success in educating the nation's rapidly growing ELL population, and inadequate second language acquisition of native English speakers. The need for improvement in these areas is great and their importance grows daily. Further studies of preschool TWI can help to produce and refine practical solutions to these educational problems.

References

- August, D. & Hakuta, K. (1997). *Improving schooling for language- minority children: A research agenda*. Washington, DC: National Research Council.
- Barnett, W. S., Hustedt, J. T., Robin, K. B., & Schulman, K. L. (2004). *The state of preschool: 2004 state preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.
- Bialystok, E. & Herman, H. (1999). Does bilingualism matter for early literacy? *Bilingualism: Language and Cognition*, 2(1), 35-44.
- Bialystok, E., Majumder, S., & Martin, M. (2003). Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics*, 24, 27-44.
- Campbell, R. & Sais, E. (1995). Accelerated metalinguistic (phonological) awareness in bilingual children. *British Journal of Developmental Psychology*, 13, 61-68.
- Campos, S. (1991). The Carpinteria Preschool Program: A long-term effects study. In E. Garcia and B. McLaughlin (Eds.), *Meeting the challenge of linguistic and cultural diversity in early education*. New York: Teachers' College.
- Collins, R. & Ribeiro, R. (2004). Toward an early care and education agenda for Hispanic children. *Early Childhood Research & Practice*, 6(2). Retrieved April 14, 2005 from <http://ecrp.uiuc.edu/v6n2/collins.html> .
- Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In California State Department of Education, (Ed.), *Schooling and language minority students: A theoretical framework* (pp. 3-49). Los Angeles, CA: California State University: Evaluation, Dissemination and Assessment Center.
- Diaz, R., Padilla, K., & Weathersby, E. (1991). The effects of bilingualism on preschoolers' private speech. *Early Childhood Research Quarterly*, 6, 377-393.
- Dunn, L. & Dunn, L. (1997). *Peabody Picture Vocabulary Test, Third edition (PPVT-III)*. Circle Pines, MN: American Guidance Service.
- Dunn, L., Lugo, D., Padilla, E., & Dunn, L. (1986). *Test de Vocabulario en Imagenes Peabody*. Circle Pines, MN: American Guidance Service.
- English, J.P., Leafstedt, J., Gerber, M. M., & Villaruz, J. (2001, April). *Individual differences in phonological skills for Spanish speaking kindergartners learning English: Relationship between English and Spanish phonological measures*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

Garcia, G. (2000). Bilingual children's reading. In M. Kamil, P. Mosenthal, P. Pearson, & R. Barr (Eds.) *Handbook of reading research, Vol III* (pp. 813-834). Mahwah, NJ: Lawrence Erlbaum.

Gormley, W., Gayer, T., Phillips, D., & Dawson, B. (2004). *The effects of universal pre-k on cognitive development* (Working paper #4). Washington, DC: Georgetown University.

Greene, J.P. (1997). A meta-analysis of the Rossell and Baker review of bilingual education research. *Bilingual Research Journal*, 21 (2/3).

Hukata, K. (1986). *Mirror of language: The debate of bilingualism*. New York, NY: Basic Books.

Harms, T., & Clifford, R., & Cryer, D. (1998). *Early Childhood Environmental Rating Scale- Revised*. New York: Teacher's College Press.

Hohmann, M. & Weikart, D.P. (1995). *Educating young children: Active learning practices for preschool and child care programs*. Ypsilanti, MI: High/Scope Press.
Howard, E. R., Sugarman, J. & Christian, D. (2003). *Trends in two-way immersion education: A review of the research* (Report 63). Baltimore, MD: CRESPAR/Johns Hopkins University.

Howard, E. R., Sugarman, J. & Christian, D. (2003). *Trends in two-way immersion education: A review of the research* (Report 63). Baltimore, MD: CRESPAR/Johns Hopkins University.

Johnson, D. & Walker, T. (1991). A follow-up evaluation of the Houston Parent-Child Development Center: School performance. *Journal of Early Intervention*, 15(3), 226-236.

Kester, E. S. & Pena, E. D. (2002). Limitations of current language testing practices for bilinguals. *Eric Digest*. (Eric Identification: ED470203).

Lamy, C. (2004, April). *Measuring early literacy: assessing children and classrooms to improve practice*. Presentation at the International Reading Association's annual conference, Reading Research, Reno, NV.

Lindholm-Leary, K. (2005). *Review of research and best practices on effective features of dual language education programs*. San Jose, CA: San Jose State University.

Lopez, L. M. (2005). *Following the development of phonological awareness skills in low-income bilingual children from pre-k through kindergarten*. Presentation at the annual Society for Research in Child Development (SRCD) conference, Atlanta, GA.

Lopez, L. M. & Greenfield, D. B. (2004). The cross-language transfer of phonological skills of Hispanic Head Start children. *Bilingual Research Journal*, 28(1), 1-18.

NCEDL (2005). Pre-kindergarten in eleven states: NCEDL's multi-state study of prekindergarten & study of state-wide early education programs (SWEEP). Preliminary descriptive report. Retrieved August 5, 2005 from http://www.fpg.unc.edu/~ncedl/pages/products.cfm#sweep_ms .

NIEER (2005). *Support for English language learners classroom assessment*. New Brunswick, NJ: National Institute for Early Education Research (NIEER).

Pearson, B. Z., Fernandez, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, 43(1), 93-120.

Peisner-Feinberg, E. & Burchinal, M. (1997). Relations between preschool children's child care experiences and concurrent development: The cost, quality, and outcomes study. *Merrill-Palmer Quarterly*, 43 (3), 451-477.

Puma, M., Bell, S., Cook, R., Heid, C., Lopez, M., Zill, N., Shapiro, G., Broene, P., Mekos, D., Rohacek, M., Quinn, L., Adams, G., Freidman, J. & Bernstein, H. (2005). *Head Start impact study: First year findings*. Washington, DC: US Department of Health and Human Services, Administration for Children and Families.

Restrapo, M. & Siverman, S. (2001). Validity of the Spanish preschool language scale-3 for use with bilingual children. *American Journal of Speech Language Pathology*, 10, 382-393.

Reynolds, A. J. (1991). *Bilingualism, multiculturalism, and second language learning*. Hillsdale, NJ: Erlbaum.

Riccio, C. A., Amado, A., Jimenez, S., Hasbrouck, J. E., Imhoff, B., & Denton, C. (2001). Cross-linguistic transfer of phonological processing: Development of a measure of phonological processing in Spanish. *Bilingual Research Journal*, 25(4), [Electronic version].

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

Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In B. K. Shapiro, P. J. Accardo & A. J. Capute (Eds.). *Specific reading disability: A view of the spectrum*. (pp. 77-121). Timonium, MD: York Press.

- Short, D. (1993). *Integrating language and culture in middle school American history classes* (Educational Practice Rep. No. 8). Santa Cruz, CA and Washington, DC: National Center for Research on Cultural Diversity and Second Language Learning.
- Slavin, R. E. & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Educational Research*, 75(2), 247-281.
- Smith S., Davidson, S., Weisenfeld, G. & Katsaros, S. (2001). *Supports for Early Literacy Assessment (SELA)*. New York: New York University.
- Smith, T, Kleiner A., Parsad, B., & Farris, E. (2003). *Prekindergarten in U. S. public schools: 2000-2001 (NCES 2003-019)*. Washington, DC: US Department of Education, National Center for Educational Statistics. ED 473 692.
- Snow, C., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Sylva, K. & Siraj-Blatchford, I. (2001, July). *The relationship between children's developmental progress in the pre-school period and two rating scales*. Paper presented at the International ECERS Network Workshop, Santiago, Chile.
- Tabors, P. O., Paez, M. M., & Lopez, L. M. (2003). Dual language abilities of bilingual four-year olds: Initial findings from the early childhood study of language and literacy development of Spanish-speaking children. *NABE Journal of Research and Practice*, 1(1), 70-91.
- Tabors, P. O. & Snow, C. E. (2001). Young bilingual children and early literacy development. In S. B. Neuman & D. K. Dickinson *Handbook of early literacy* (pp. 159-178). New York, NY: Guilford Press.
- U. S. Census Bureau (2000). *Census 2000 summary file3 (SF 3): Tables P145A-P145I*. Retrieved April 15, 2005 from http://factfinder.census.gov/servlet/DatasetTableListServlet?_ds_name=DEC_2000_SF3_U&_type=table&lang=en&_program=DEC&ts=7564142654
- Whitehurst, G. J. & Lonigan, C. J. (2001). Emergent literacy: Development from prereaders to readers. In S. B. Neuman & D. K. Dickinson (Eds.) *Handbook of early literacy research* (pp. 11-29). New York, NY: Guilford Press.
- Willig, A. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research*, 55, 269-317.
- Winsler, A., Diaz, R., Espinosa, L., & Rodriguez, J. (1999). When learning a second language does not mean losing the first: Bilingual language development in low-income,

Spanish-speaking children attending bilingual preschool. *Child Development*, 70(2), 349-362.

Woodcock, R. & Johnson, M. (1989) *Woodcock-Johnson Psycho-Educational Battery-Revised*. Allen, TX: TWIM Teaching Resources.

Woodcock, R. & Munoz-Sandoval, A.F. (1996). *Bateria Woodcock-Munoz: Pruebas de habilidad cognitive-Revisada*. Chicago, IL: Riverside Publishing.

Wong Fillmore, L. (1991). When learning a second language means losing the first. *Early Childhood Research Quarterly*, 6, 323-346.

Table 1.

Sample Description

| | Two-Way Immersion | | English Immersion | |
|-----------------------|-------------------|---------|-------------------|---------|
| | Number | Percent | Number | Percent |
| | 85 | 100 | 60 | 100 |
| Gender | | | | |
| Male | 35 | 41.2 | 32 | 53.3 |
| Female | 50 | 58.8 | 28 | 46.7 |
| Child's Age | | | | |
| 3 years old | 40 | 47.1 | 22 | 36.7 |
| 4 years old | 45 | 52.9 | 38 | 63.3 |
| Home Language | | | | |
| Spanish | 49 | 57.6 | 35 | 58.3 |
| English | 30 | 35.3 | 21 | 35.0 |
| Other | 6 | 7.1 | 4 | 6.7 |
| Race-Ethnicity | | | | |
| White/non-Hispanic | 6 | 7.1 | 5 | 8.3 |
| African American | 9 | 10.6 | 8 | 13.3 |
| Hispanic | 67 | 78.8 | 45 | 75.0 |
| Other | 3 | 3.5 | 2 | 3.3 |

Table 2.

Observations of Classroom Quality

| | Two-Way Immersion Classrooms | | | English Immersion Classrooms | | | Total | | |
|------------|------------------------------|------|-----------|------------------------------|-------|-----------|-------|------|-----------|
| | N | Mean | Std. Dev. | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| ECERS- R | 20 | 5.11 | .63 | 16 | 4.86 | .65 | 36 | 5.0 | .64 |
| SELA | 20 | 3.23 | .42 | 16 | 3.01 | .53 | 36 | 3.13 | .48 |
| SELLCA | | | | | | | | | |
| L1 Use | 10 | 3.90 | .32 | 16 | 1.81* | .98 | | | |
| Culture | 10 | 3.50 | .97 | 16 | 2.25* | .58 | | | |
| L2 Support | 10 | 3.70 | .48 | 16 | 3.25 | .68 | | | |

ECERS-R=Early Childhood Environmental Rating Scale

SELA=Supports for Early Literacy Assessment

SELLCA=Supports for English Language Learners Classroom Assessment (Use of Primary Language,

*t-test, $p < .05$

Table 3.

Regression analysis estimating TWI effects on English language measures (all children), controlling for primary language, age, and pretest score.

| Variable | <u>PPVT (n=131)</u> | | <u>WCJ-PV (n=131)</u> | | <u>WCJ-AP (n=128)</u> | | <u>PD-E (n=130)</u> | | <u>Letter ID-E (n=131)</u> | | <u>Rhyme-E (n=131)</u> | |
|----------------|---------------------|------|-----------------------|------|-----------------------|------|---------------------|------|----------------------------|------|------------------------|------|
| | B | S.E. | B | S.E. | B | S.E. | B | S.E. | B | S.E. | B | S.E. |
| Constant | 3.70 | 6.74 | 8.57 ** | 1.40 | .66 | 3.78 | .59 | 2.21 | 3.70 | 4.56 | -4.25 ** | 1.39 |
| TWI v. EI | 1.34 | 1.67 | -.20 | .37 | 1.21 | .90 | 1.07 + | .56 | -.32 | 1.16 | .00 | .35 |
| Spanish L1 | -2.10 | 1.85 | -.43 | .41 | .13 | .90 | .15 | .56 | -.29 | 1.19 | .07 | .35 |
| Age | .33 * | .16 | .04 | .03 | .09 | .09 | .18 ** | .05 | .09 | .09 | .11 ** | .03 |
| Pretest Score | .72 ** | .07 | .54 ** | .04 | .67 ** | .14 | .15 * | .07 | .06 ** | .10 | .58 ** | .10 |
| R ² | .70 | | .71 | | .31 | | .24 | | .50 | | .39 | |
| F | 72.71 | | 77.38 | | 13.59 | | 9.65 | | 31.32 | | 19.84 | |

+ p<.10

* p<.05

** p<.01

All F- values were significant at p<.001.

PPVT- The Peabody Picture Vocabulary Test

WCJ-PV- Woodcock Johnson Picture Vocabulary

WCJ-AP- Woodcock Johnson Applied Problems

PD-E- Phoneme Deletion- English

Letter ID-E- Alphabet Recognition- English

Rhyme- E- Rhyme Recognition- English

Table 4.

Regression analysis estimating TWI program effects on Spanish language measures (Spanish dominant children only), controlling for age and pretest score.

| Variable | <u>TVIP (n=74)</u> | | <u>WCM-PV (n=74)</u> | | <u>WCM-AP (n=74)</u> | | <u>PD-S (n=74)</u> | | <u>Letter ID-S (n=74)</u> | | <u>Rhyme-S (n=74)</u> | |
|----------------|--------------------|------|----------------------|------|----------------------|------|--------------------|------|---------------------------|------|-----------------------|------|
| | B | S.E. | B | S.E. | B | S.E. | B | S.E. | B | S.E. | B | S.E. |
| Constant | 10.66 | 7.71 | 5.69 * | 2.74 | -4.96 * | 2.43 | 2.04 | 2.45 | -5.67 | 5.83 | -3.24 | 2.35 |
| TWI v. EI | 6.39 ** | 1.87 | .31 | .69 | .24 | .61 | .79 | .61 | -3.01 * | 1.47 | 1.18 * | .57 |
| Age | -.14 | .16 | -.06 | .06 | .27 ** | .05 | .15 ** | .05 | .27 * | .12 | .09 + | .05 |
| Pretest Score | .94 ** | .10 | .74 ** | .10 | .03 | .03 | .24 ** | .07 | .90 ** | .14 | .53 ** | .15 |
| R ² | .60 | | .45 | | .34 | | .33 | | .47 | | .27 | |
| F | 35.48 | | 19.41 | | 11.75 | | 11.23 | | 20.32 | | 8.64 | |

+ p<.10

* p<.05

** p<.01

All F- values were significant at p<.001.

TVIP- Test de Vocabulario en Imagenes Peabody

WCM-PV- Woodcock-Muñoz Picture Vocabulary

WCM-AP- Woodcock-Muñoz Applied Problems

PD-S- Phoneme Deletion- Spanish

Letter ID-S- Alphabet Recognition- Spanish

Rhyme-S- Rhyme Recognition- Spanish