## Abstract Title Page

*Not included in page count.*

**Title:**
**Immersion versus primary language effects for growth in Spanish and English letter-word identification among children, classrooms, and schools**

**Author(s):**

Lee Branum-Martin. Texas Institute for Measurement, Evaluation, and Statistics, University of Houston. Lee.Branum-Martin@times.uh.edu

Paras D. Mehta. Texas Institute for Measurement, Evaluation, and Statistics, University of Houston. Paras.Mehta@times.uh.edu

David J. Francis. Texas Institute for Measurement, Evaluation, and Statistics, University of Houston. dfrancis@uh.edu

## Abstract Body

*Limit 5 pages single spaced.*

### Background / Context:

Investigations of bilingual programs pose intense analytic challenges to appropriate inference, including longitudinal outcomes, school differences, and program differences delivered at the classroom level. Recent research syntheses suggest that native language instruction helps Spanish speaking children develop early literacy skills in English (Francis, Lesaux, & August, 2006; Greene, 1998; Rolstad, Mahoney, & Glass, 2005; Slavin & Cheung, 2005; Willig, 1985). Native language instruction may not only be beneficial for the learning of English, may also promote the learning of Spanish (Branum-Martin, Foorman, Francis, & Mehta, 2010). However, classrooms may differ widely in instruction, even under the same program labels (Branum-Martin, et al., 2010; Branum-Martin et al., 2006; Branum-Martin et al., 2009; Cirino, Pollard-Durodola, Foorman, Carlson, & Francis, 2007; Foorman, Goldenberg, Carlson, Saunders, & Pollard-Durodola, 2004; Saunders, Foorman, & Carlson, 2006). While randomized studies may be helpful for causal inference, assigning language-minority students to English versus Spanish instruction may be exceedingly difficult to accomplish in dependable numbers of classrooms and schools. Specifically, we wish to evaluate a 4-level cross classified structure of longitudinal measures within students, students cross-classified by teachers, and teachers nested within schools.

### Purpose / Objective / Research Question / Focus of Study:

The current project estimates program effects (English immersion versus primary language instruction using varying degrees of Spanish) upon English and Spanish word identification in the context of changing classrooms (cross-classification) from kindergarten through second grade. Letter-word identification is an important predictor of early reading achievement (Scarborough, 2001, 2005). In particular, we are interested in classroom and school differences not merely as noise to be removed, but as indicators of substantively important differences in implementation and practice.

### Setting:

All schools participating in the completed quasi-experimental study met the following criteria: (a) at least acceptable levels of achievement as gauged by state accountability measures and criteria; (b) at least 40% of the school's enrollment was Hispanic; (c) at least 30% of students in Kindergarten and first grade were English language learners, (d) the school had one or more designated programs for English learners (English Immersion or a specific bilingual program model); and (e) school and district administrators consented to participate in the study. From this pool, 34 schools in urban Texas, border Texas, and urban California were selected for the final sample to maximize classroom-level representation across three major program models: English immersion, transitional, and dual language. The schools thus represent adequately performing largely Hispanic schools with instruction intended for Spanish speaking students.

### Population / Participants / Subjects:

The student participants were 1,991 children (49% female) taught by 421 teachers across the 3 years of the longitudinal project, kindergarten to second grade. Seventy-five percent of the families reported earning less than $30,000 per year, and 79% had foreign-born fathers.

**Intervention / Program / Practice:**

Trained research staff who were former bilingual teachers categorized each classroom into the type of instructional program used. For the purpose of the current study, we compare English immersion with other programs which use Spanish during instruction, labeled primary language instruction.

**Research Design:**

The data are from a quasi-experimental study from kindergarten through second grade. Students were tested in Fall and Spring of each year (September or October, and April or May, respectively). Children therefore had a total possible 6 time points. Table 1 shows the counts of teachers, counts of students, and the mean and SD for Spanish and English letter-word identification (described below) for each language program in each year of the study.

**Data Collection and Analysis:**

Children were individually tested by bilingual examiners using the Woodcock Language Proficiency Battery, Revised (Woodcock, 1991; Woodcock & Muñoz-Sandoval, 1995). The outcomes used in the present study were the English and Spanish versions of the Letter-Word Identification subtests in which students named letters and read words of increasing difficulty. The Rasch-scaled W-scores were used to index growth from kindergarten through second grade. Students were tested in the fall and spring of each year, yielding a maximum of 6 observations per child.

These 6 time points were nested within children, children were nested within teachers, and teachers were nested within schools. Children and schools likely grow at different rates over 6 time points in 3 years, but students change classrooms each year, resulting in a cross-classified nesting of students by classroom, with classrooms constituting the grouping of teachers with different groups of students. This structure of longitudinal outcomes, children, classrooms, and schools yields a 4-level cross-classified modeling problem.

Exploratory graphs of student and school level trajectories suggested that a linear or quadratic model for student or school trajectories would be adequate. The result was a 4-level model of growth for the English and Spanish outcomes separately. Because students changed teachers each year and there were two observations each year, random variability in fall and spring classroom effects was modeled. The general form of the model was:

Level 1: outcome for *time t, student i, class j, school k*

$$Y_{tijk} = \pi_{0\,ijk} + \pi_{1\,ijk} * \text{time}_t + \pi_{2\,ijk} * \text{time}^2_t + u_{t\bullet jk} + e_{ijk}$$

Level 2: student (*i*)

$\pi_{0\,ijk} = \beta_{0\,jk} + r_{0\,ijk}$ student intercept + deviation

$\pi_{1\,ijk} = \beta_{1\,jk} + r_{1\,ijk}$ student linear slope + deviation

| | |
|---|---|
| $\pi_{2\,ijk} = \beta_{2\,jk} + r_{2\,ijk}$ | student quadratic slope + deviation |

Level 3: classroom ($j$)

| | |
|---|---|
| $\beta_{0\,jk} = \gamma_{00\,k} + \gamma_{01\,k}$ * $\text{Immersion}_j$ | placeholder for intercept |
| $\beta_{1\,jk} = \gamma_{10\,k} + \gamma_{11\,k}$ * $\text{Immersion}_j$ | placeholder for linear slope |
| $\beta_{2\,jk} = \gamma_{20\,k} + \gamma_{21\,k}$ * $\text{Immersion}_j$ | placeholder for quadratic effect |
| $u_{t\bullet jk}$ | classroom disturbance at time $t$ |

Level 4: school ($k$)

| | |
|---|---|
| $\gamma_{00\,k} = \delta_{00} + w_{0\,k}$ | intercept + school deviation |
| $\gamma_{01\,k} = \delta_{01}$ | immersion effect for intercept |
| $\gamma_{10\,k} = \delta_{10} + w_{1\,k}$ | linear slope + school deviation |
| $\gamma_{11\,k} = \delta_{11}$ | immersion effect for linear slope |
| $\gamma_{20\,k} = \delta_{20} + w_{2\,k}$ | quadratic slope + school deviation |
| $\gamma_{21\,k} = \delta_{21}$ | immersion effect for quadratic |

At level 1, the equation represents an outcome at time $t$ as a function of average status ($\pi_{0ijk}$) plus a linear slope ($\pi_{1\,ijk}$) and measurement error ($e_{ijk}$), distributed normally, identically, and independently across time.

Level 2 represents student deviations in intercept ($r_{0\,ijk}$) and slope ($r_{1\,ijk}$) and quadratic slope ($r_{2\,ijk}$). Student random effects are allowed to covary.

At level 3, we estimate the average program effect for being in an Immersion classroom ($\gamma_{01\,k}$), the interaction of program and slope ($\gamma_{11\,k}$), or the differential slope due to program. Because slopes are not identified for classrooms with only two time points, classroom deviations from the average trajectory are represented by a time-specific random effect ($u_{t\bullet jk}$). With two waves of measurement in each year, there are two possible deviations per classroom per grade. These deviations in fall and spring of each year may covary within year.

At level 4, schools vary in their level of performance (intercepts deviate $w_{0\,k}$ about the grand intercept, $\delta_{00}$), their linear slopes ($w_{1\,k}$), and possibly their quadratic slopes ($w_{2\,k}$). The school random effects covary. Average program effects for intercept ($\delta_{01}$), linear slope ($\delta_{11}$), and quadratic slope ($\delta_{21}$) are fixed at the school level.

The random effects just presented at the student, classroom, and school levels may have a completely free structure as suggested in the preceding presentation, they may have a simpler structure, or they may differ across instructional programs. For example, schools may not differ much in the curvature of their growth trajectories, so a model constraining the $w_{2\,k}$ parameter to zero and dropping its covariances may be a more parsimonious model.

At the classroom level, spring performance may correlate highly with fall performance and have higher variability. A more parsimonious model may be one of incremental variance where spring variability is an additional deviation above fall variability. Such an incremental model estimates two variances without a covariance. In addition, classroom variability may differ between the two programs. Models of separate classroom effects will be tested.

Finally, at the student level, it is possible that instructional programs have different covariance structures. That is, in different programs, the relation of the growth factors may be different for the students. While the fixed effects capture differences in average trajectory between programs (the $\delta$ terms), the variability in and relations among intercept and slopes may differ across programs. For example, it is possible that the amount of quadratic curvature among students may be different across the two programs.

**Findings / Results:**

Figure 1 shows the density of scores for each language program in each time point for English and Spanish letter-word identification W-scores. The top row of Figure 1 shows each semester for English, while the bottom row shows the program distribution of scores in each semester for Spanish. Each cell in the figure represents a single semester, like a histogram, with blue representing Spanish maintenance programs and red representing English immersion programs. In each cell, a vertical dotted line represents the grade-level expected referenced W-score for children of that age in that semester (Woodcock, 1991; Woodcock & Muñoz-Sandoval, 1995). Figure 1 provides additional distributional information not present in Table 1.

Table 2 presents the fit indices of the models for English and Spanish letter-word identification. Each line of the table represents a model with different formulations of random effects for students, classrooms, and schools. The first column of Table 2 shows the student random effects. The structures fit were linear, quadratic, and quadratic by program. The quadratic by program structure indicates that all three random effects (intercept, linear, and quadratic) were free to differ in their variances and covariances between programs. All models were fit using REML estimation in the *lmer* function from the *lme4* package in *R 2.12.1* (R Development Core Team, 2010).

The second column of Table 2 shows the covariance structures for classrooms for that model. If a modification from an unconstrained quadratic model was made at the school level, it is noted in parentheses. Three types of structures were fit to classrooms: unstructured, incremental, and yearly by program. The unstructured random effects represent separate fall and spring variances with a freely estimated covariance for each year (3 parameters per year). Incremental variance estimates spring variance as an increased deviation over fall variance (2 parameters per year). The yearly effects estimates a single classroom intercept per year (1 parameter).

The models in the rows of Table 2 are sorted from worst to best in terms of their fit by the Bayesian Information Criterion (BIC) and loglikelihood (LL). So the first line of Table 2 represents a linear random effects student model with incremental fall-spring variances for classrooms and linear random effects for schools. The second line shows the second-worst fitting model had separate student random effects by program and yearly classroom intercepts by program. The yearly intercept structure did not fit well in other models, so no other versions are reported. Three of the models for English letter-word identification resulted in some of the classroom variances being restricted to zero ("zeros" in the note column) during estimation in the *lmer* function in *R*. Such estimation problems likely indicate over-parameterization, so these models are reported, but not considered valid.

The English results in the top half of Table 2 show that the best model was the second from the bottom: quadratic random effects for students separately by program, incremental classroom variances by program, and linear school random effects.

The lower half of Table 2 shows the results for the Spanish models. No classroom variances estimated at zero. Coincidentally, the same structure of random effects fit for Spanish as English: quadratic random effects for students separately by program, incremental classroom variances by program, and linear school random effects.

Figure 2 presents this resulting model as a SEM diagram (modified after McArdle, 1988). The parameters from the equations listed before are labeled. Representing the equations this way shows how a multilevel regression model, even with a single outcome, is actually a model of covariances. At the bottom of Figure 2, the common residual, $e_{ijk}$, shows the restriction of equal variances and zero covariances. At the student level, the random effects for the intercept ($r_{0\ ijk}$), linear ($r_{1\ ijk}$) and quadratic growth ($r_{2\ ijk}$) are shown separately for each program. Their respective variances ($\tau_{\pi}$) are shown, and the covariances are left unlabelled, but are freely estimated (see correlations in Table 3 below).

At the classroom level in Figure 2, the effect for Immersion program is shown as a deviation to the growth parameters. Because each classroom only had two time points per year, variability in growth parameters is not identified. Instead, classroom variability is modeled separately for each semester. The resulting classroom structure was incremental: spring variability was modeled as an increase over fall variability. The gray linking lines show these relations. Classroom variability was also modeled separately for each program, so these random effects are duplicated in the diagram.

Lastly, schools were estimated to differ in average performance (intercept) and linear slope. School variability is shown for intercept ($\tau_{\gamma 11}$) and linear slope ($\tau_{\gamma 22}$). The covariance is estimated also (see curved arrow). Paths from the triangle at the top shows the overall fixed effects for the Primary language program (Immersion deviations appear at the classroom level). Overall, Figure 2 helps to visualize the covariances not easily seen in the scalar equations.

The specific results of these models are shown in Table 3. Fixed effects are at the top and random effects appear below. Time for the six waves of the study was centered at the middle of first grade (values = -2.5, -1.5, -0.5, 0.5, 1.5, 2.5) so that each unit of growth represents one year of school. The Immersion variable was coded 1, so that Immersion estimates represent deviations from the Primary language program.

The intercept at the top of Table 3 represents a model-predicted average of 416.6 at the middle of first grade (time zero) in Primary language instruction. The linear value of indicates that children in Primary language instruction could be expected to grow 22.3 W-score units per year, with a downward curvature (quadratic term) of -0.2 W-score units per year.

The fixed effects for English letter-word identification indicate that Immersion programs on average would be expected to score 20.7 points higher than Primary programs at the middle of first grade. Immersion programs had a linear slope -4.0 points lower than the given 22.3, and a steeper quadratic curve at -1.8 more than the -0.2 than the Primary program. The overall profiles of these curves will be shown graphically.

The random effects in Table 3 are shown for the four levels: student, classroom, school, and residual. For each level, the variance, SD, and correlations (if relevant) are shown. In English, the student random effects show greater variability in the Primary program than in the Immersion program. At the classroom level in English, the incremental variance results are shown. Each spring estimate is an amount above the variance shown for fall. Thus, while the variance for fall in kindergarten was 4.7, the variance for spring would be 55.4 (4.7 + 50.7), yielding a spring total SD of 7.4 [sqrt(55.4)]. It should be noted that the classroom variance for Immersion in the spring of first grade was very small ($2 \times 10^{-6}$) but was not restricted to zero.

At the school level, random effects were only fit for the intercept and linear slope. Schools differed 6.1 units (±1 SD) in average performance and 1.9 units in their slopes. The standard deviation at the student level was 12.5 W-score units.

The right-hand side of Table 3 shows the results for Spanish letter-word identification. At the middle of first grade, students in Primary language instruction scored on average 473.8 W-score units, grew at 25.7 units per year, with a -6.3 quadratic effect per year. There were strong differences for Immersion programs' performance in Spanish. Children could be expected to perform 72.9 points lower, have a 11.5 unit lower slope, and 5.5 units less of a quadratic bend in their trajectory.

The student random effects were different between Primary and Immersion programs, with Primay programs having somewhat higher variability, but the SD were fairly close. At the classroom level in Spanish, the Primary language programs had much higher variability than did the Immersion classrooms. Schools differed on average 10.0 units in average performance (SD) and 4.7 units (SD) in linear slope.

Figure 3 shows the model-predicted trajectory for each program ±1 classroom SD for English letter-word identification. Figure 4 shows the trajectories for each program for Spanish. These graphs allow visual inspection of the results of the program-specific fixed effects, as well as the relative influence of classroom variability which is specific to program and semester.

**Conclusions:**

Substantively, the current results suggest that in English, Spanish language instruction may facilitate the learning of English over time, as evidenced by the maintenance programs catching up to the English performance of the immersion programs (Figure 2). In Spanish letter-word identification, immersion programs show improving performance over time, but at a slower pace than the Spanish maintenance programs. These program-by-time interactions, however, are not completely correct and will require further verification.

The model here is complex and preliminary, so these conclusions must be viewed in the light of substantial limitations. First, this is a quasi-experiment, so important and perhaps pre-existing differences may be present at the student and school/community levels. In addition, classroom instruction is known to differ, even within these program types, so classroom level measures of instruction may help to clarify these results. These models will be extended to include child and classroom covariates.

Second, missing data requires further examination. We have applied quadratic growth models to the student level with effects for patterns of missing data (e.g., early dropouts, newcomers, and longitudinal participants). These preliminary models suggest no serious problems, but further investigation is warranted.

Third, the current analysis is univariate in the outcomes: one language at a time. In future analyses, we hope to extend these models to examine cross-language effects, such as the influence of Spanish performance upon later English performance.

So far, the descriptive statistics and density plots in Figure 1 suggest promise for these preliminary models. The results support one of the central tenets of bilingual education: that native language instruction may foster growth in English achievement. Viewed from the opposite direction, however, Spanish instruction does not appear to convey special advantage in English letter-word identification in these early grades. Performance in Spanish seems to require sustained instruction—that is, growth in Spanish early literacy skills may slow under English immersion.

More importantly, however, is that the current approach provides empirical estimates of school and classroom variability. Often, classroom variability was comparable to school variability, and this variability could equal half a semester worth of growth. Teacher and school variability therefore may have strong implications for our evaluation of language programs for Spanish speakers as well as for the development of literacy skills within and across language.

## Appendices

*Not included in page count.*

### Appendix A. References


Branum-Martin, L., Foorman, B. R., Francis, D. J., & Mehta, P. D. (2010). Contextual effects of bilingual programs on beginning reading. *Journal of Educational Psychology, 102*(2), 341-355.

Branum-Martin, L., Mehta, P. D., Fletcher, J. M., Carlson, C. D., Ortiz, A., Carlo, M. S., et al. (2006). Bilingual phonological awareness: Multilevel construct validation among Spanish-speaking kindergarteners in transitional bilingual education classrooms. *Journal of Educational Psychology, 98*(1), 170-181.

Branum-Martin, L., Mehta, P. D., Francis, D. J., Foorman, B. R., Cirino, P. T., Miller, J. F., et al. (2009). Pictures and words: Spanish and English vocabulary in classrooms. *Journal of Educational Psychology, 101*(4), 897-911.

Cirino, P. T., Pollard-Durodola, S. D., Foorman, B. M., Carlson, C. D., & Francis, D. J. (2007). Teacher characteristics, classroom instruction, and student literacy and language outcomes in bilingual kindergartners *Elementary School Journal, 107*(4), 341-364.

Foorman, B. R., Goldenberg, C., Carlson, C. D., Saunders, W., & Pollard-Durodola, S. D. (2004). How teachers allocate time during literacy instruction in primary-grade English language learner classrooms. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 289-328). Baltimore, MD: Brookes.

Francis, D. J., Lesaux, N., & August, D. (2006). Language of instruction. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the national literacy panel on language-minority children and youth* (pp. 365-413). Mahwah, NJ: Erlbaum.

Greene, J. P. (1998). *A meta-analysis of the effectiveness of bilingual education*. Claremont, CA: Tomas Rivera Policy Institute.

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 561-614). New York: Plenum Press.

R Development Core Team. (2010). R: A language and environment for statistical computing (Version 2.12.1). Vienna, Austria: R Foundation for Statistical Computing.

Rolstad, K., Mahoney, K., & Glass, G. V. (2005). The big picture: A meta-analysis of program effectiveness research on English language learners. *Educational Policy, 19*(4), 572-594.

Saunders, W. M., Foorman, B. R., & Carlson, C. D. (2006). Is a separate block of time for oral English language development in programs for English learners needed? . *Elementary School Journal, 107*(2), 181-198.

Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In D. K. Dickinson & S. B. Neuman (Eds.), *Handbook of early literacy research* (Vol. 1, pp. 97-110). New York: Guilford Press.

Scarborough, H. S. (2005). Developmental relationships between language and reading: Reconciling a beautiful hypothesis with some ugly facts. In H. W. Catts & A. G. Kamhi (Eds.), *The connections between language and reading disabilities* (pp. 3-24). Mahwah, NJ: Erlbaum.

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), 247-284.

Willig, A. C. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research, 55*(3), 269-317.

Woodcock, R. W. (1991). *Woodcock Language Proficiency Battery-Revised*. Itasca, IL: Riverside.

Woodcock, R. W., & Muñoz-Sandoval, A. F. (1995). *Woodcock Language Proficiency Battery-Revised, Spanish Form*. Chicago: Riverside.

## Appendix B. Tables and Figures

*Not included in page count.*

*Table 1: Descriptive Statistics for Letter-Word Identification W-scores*

| | | Kindergarten | | First Grade | | Second Grade | |
|---|---|---|---|---|---|---|---|
| Statistic | Program | Fall | Spring | Fall | Spring | Fall | Spring |
| Teachers, *n* | Maintenance | 53 | 55 | 90 | 93 | 100 | 102 |
| | Immersion | 34 | 35 | 70 | 68 | 67 | 66 |
| Students, *n* | Maintenance | 892 | 812 | 873 | 865 | 848 | 799 |
| | Immersion | 417 | 390 | 426 | 480 | 456 | 398 |
| English mean | Maintenance | 366 | 391 | 408 | 435 | 458 | 473 |
| | Immersion | 382 | 414 | 431 | 453 | 466 | 481 |
| SD | Maintenance | 18 | 30 | 34 | 35 | 28 | 27 |
| | Immersion | 21 | 25 | 23 | 28 | 22 | 22 |
| Spanish mean | Maintenance | 379 | 442 | 460 | 495 | 502 | 510 |
| | Immersion | 363 | 384 | 397 | 424 | 446 | 454 |
| SD | Maintenance | 28 | 51 | 49 | 39 | 37 | 37 |
| | Immersion | 23 | 36 | 40 | 43 | 39 | 44 |

*Table 2: Fit of models for Letter-word Identification*

| Student effects | Teacher effects, school | BIC | LL | note |
|---|---|---|---|---|
| **English** | | | | |
| Linear | Incremental, linear school | 67,466 | -33,648 | |
| Quadratic by program | Yearly by program | 67,239 | -33,481 | |
| Quadratic | Unstructured | 67,002 | -33,376 | |
| Quadratic | Unstructured by program | 66,964 | -33,317 | zeros |
| Quadratic | Incremental by program | 66,956 | -33,340 | zeros |
| Quadratic | Incremental | 66,956 | -33,366 | |
| Quadratic | Incremental, linear school | 66,949 | -33,376 | |
| Quadratic by program | Incremental by program, linear school | 66,930 | -33,313 | |
| Quadratic by program | Incremental by program | 66,929 | -33,299 | zeros |
| **Spanish** | | | | |
| Linear | Incremental, linear school | 74,275 | -37,053 | |
| Quadratic by program | Yearly by program | 74,224 | -36,973 | |
| Quadratic | Unstructured | 73,904 | -36,827 | |
| Quadratic | Incremental | 73,881 | -36,829 | |
| Quadratic | Unstructured by program | 73,881 | -36,788 | |
| Quadratic | Incremental | 73,872 | -36,838 | |
| Quadratic | Incremental | 73,868 | -36,795 | |
| Quadratic by program | Incremental by program | 73,814 | -36,741 | |
| Quadratic by program | Incremental by program, linear school | 73,808 | -36,752 | |

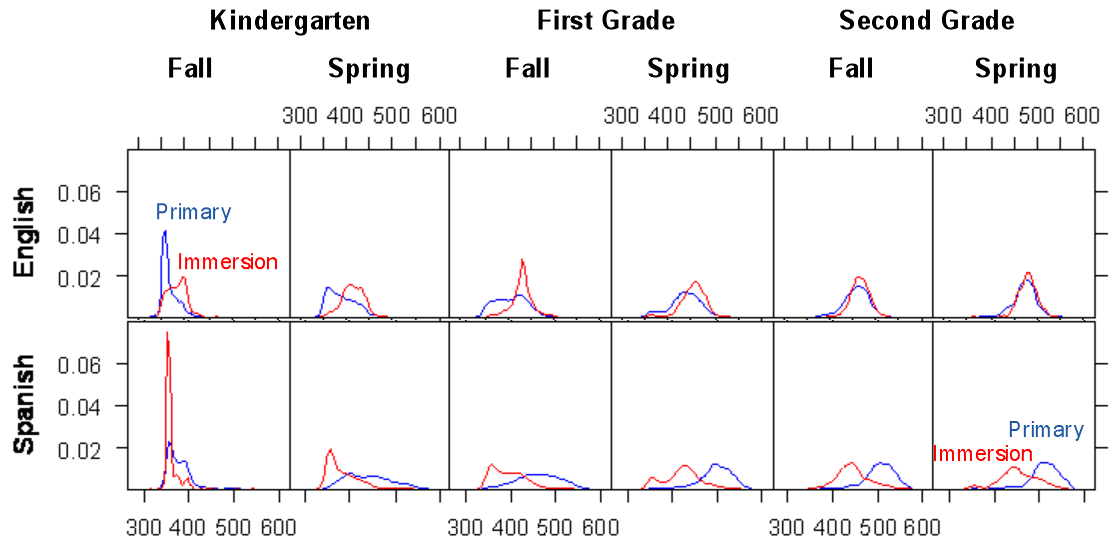
*Note: BIC = Bayesian Information Criterion. LL = Log Likelihood. Unstructured random effects represent separate variances with free covariances. Incremental classroom variance effects represent spring variance within each year as an increment over fall variance. "zeros" refers to classroom random effects which were fixed by the* lmer *program during estimation, indicating an over-parameterized model.*

*Table 3: Results for quadratic student effects by program, incremental teacher variance by program, and linear school random effects*

| | | English | | | | Spanish | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Fixed Effects** | | Estimate | SE | *t* | | Estimate | SE | *t* | |
| $\delta_{00}$ | Intercept | 416.6 | 23.1 | 18.0 | | 473.8 | 35.7 | 13.3 | |
| $\delta_{10}$ | linear | 22.3 | 1.2 | 17.8 | | 25.7 | 4.2 | 6.1 | |
| $\delta_{20}$ | quadratic | -0.2 | 1.0 | -0.2 | | -6.3 | 1.9 | -3.3 | |
| $\delta_{01}$ | Immersion | 20.7 | 37.0 | 0.6 | | -72.9 | 52.7 | -1.4 | |
| $\delta_{11}$ | linear * Imm. | -4.0 | 1.6 | -2.5 | | -11.5 | 4.5 | -2.5 | |
| $\delta_{21}$ | quad. * Imm. | -1.8 | 1.2 | -1.5 | | 5.5 | 2.2 | 2.5 | |
| **Random Effects** Level | **Program** Source | Variance | SD | Corr | | Variance | SD | Corr | |
| Student | **Primary** | | | | | | | | |
| $\tau_{\pi 11}$ | Intercept | 817.5 | 28.6 | | | 1481.9 | 38.5 | | |
| $\tau_{\pi 22}$ | linear | 14.5 | 3.8 | 0.35 | | 28.3 | 5.3 | 0.17 | |
| $\tau_{\pi 33}$ | quadratic | 5.7 | 2.4 | -0.94 | -0.09 | 8.2 | 2.9 | -1.00 | -0.17 |
| | **Immersion** | | | | | | | | |
| $\tau_{\pi 11}$ | Intercept | 524.7 | 22.9 | | | 1142.5 | 33.8 | | |
| $\tau_{\pi 22}$ | linear | 7.3 | 2.7 | 0.19 | | 31.1 | 5.6 | 0.66 | |
| $\tau_{\pi 33}$ | quadratic | 1.0 | 1.0 | -0.92 | -0.55 | 2.9 | 1.7 | -0.98 | -0.49 |
| Classroom | **Primary** | | | | | | | | |
| $\tau_{(u0)}$ | K fall | 4.7 | 2.2 | | | 15.0 | 3.9 | | |
| $\tau_{(u1)}$ | K spring | 50.7 | 7.1 | | | 366.2 | 19.1 | | |
| $\tau_{(u2)}$ | 1 fall | 50.1 | 7.1 | | | 181.6 | 13.5 | | |
| $\tau_{(u3)}$ | 1 spring | 49.3 | 7.0 | | | 213.8 | 14.6 | | |
| $\tau_{(u4)}$ | 2 fall | 2.0 | 1.4 | | | 166.2 | 12.9 | | |
| $\tau_{(u5)}$ | 2 spring | 44.9 | 6.7 | | | 69.2 | 8.3 | | |
| | **Immersion** | | | | | | | | |
| $\tau_{(u0)}$ | K fall | 13.3 | 3.6 | | | 7.4 | 2.7 | | |
| $\tau_{(u1)}$ | K spring | 44.9 | 6.7 | | | 69.8 | 8.4 | | |
| $\tau_{(u2)}$ | 1 fall | 15.7 | 4.0 | | | 52.0 | 7.2 | | |
| $\tau_{(u3)}$ | 1 spring | 0.0 | 0.0 | | | 25.7 | 5.1 | | |
| $\tau_{(u4)}$ | 2 fall | 2.2 | 1.5 | | | 6.2 | 2.5 | | |
| $\tau_{(u5)}$ | 2 spring | 19.5 | 4.4 | | | 47.6 | 6.9 | | |
| School | | | | | | | | | |
| $\tau_{\beta 11}$ | Intercept | 36.8 | 6.1 | | | 99.3 | 10.0 | | |
| $\tau_{\beta 22}$ | linear | 3.7 | 1.9 | 0.41 | | 22.2 | 4.7 | 0.62 | |
| Residual | | 155.8 | 12.5 | | | 362.3 | 19.0 | | |

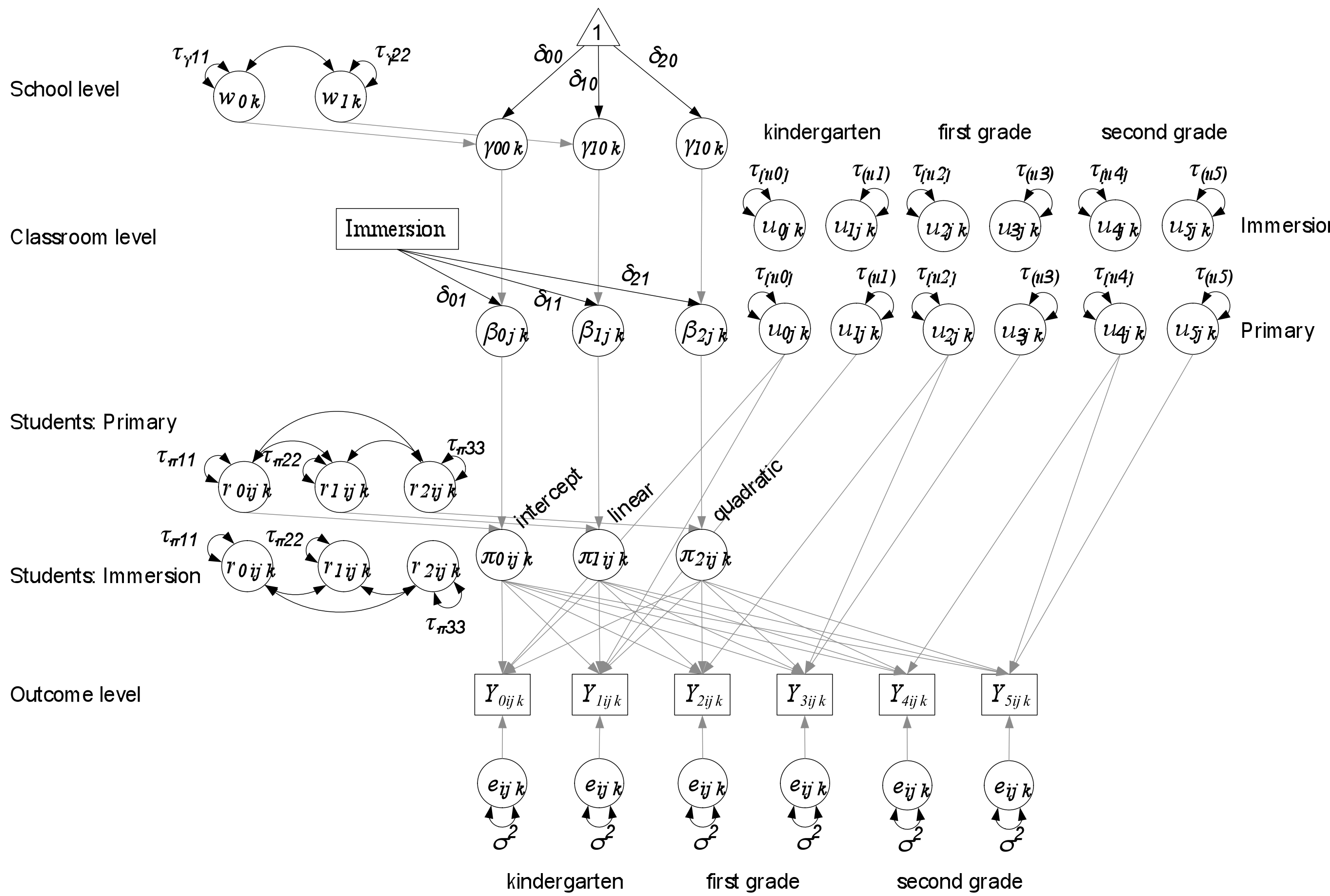
*Note*: Time was centered at the middle of the 6 time points (midyear of first grade). The variance components for classrooms represent incremental estimates within each year: spring variance is additional variance above fall. See text for details.

*Figure 1: Density of WJ-R W-scores by semester and program*




*Note. The outcomes are Rasch-based W-scores from the English and Spanish versions of the WLPB-R Letter-word Identification subtests. Solid lines represent distribution density for each program in each semester (using the default kernel smoothing in the* R *function,* densityplot*). Vertical dotted lines represent the grade-level expected reference W-score for each semester from the WLPB-R norm tables.*

*Figure 2: SEM representation of the resulting model: Quadratic random effects by program for students, incremental classroom effects by program, and linear random effects for schools.*




*Note*. Gray arrows show nesting link functions. The link lines for the growth parameters follow those for a quadratic growth model centered between time 2 and 3 (middle of first grade). Covariances for random effects are not labeled (see correlations in Table 3). To avoid visual clutter, the random effects for the Immersion program are shown beside the random effects for the Primary language program without gray link lines.

*Figure 3: English model-predicted scores for each program, with ±1 classroom SD.*

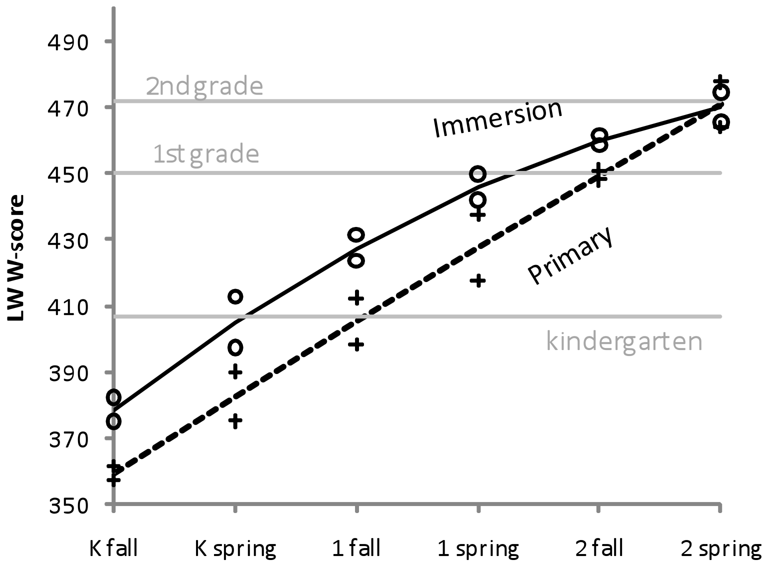
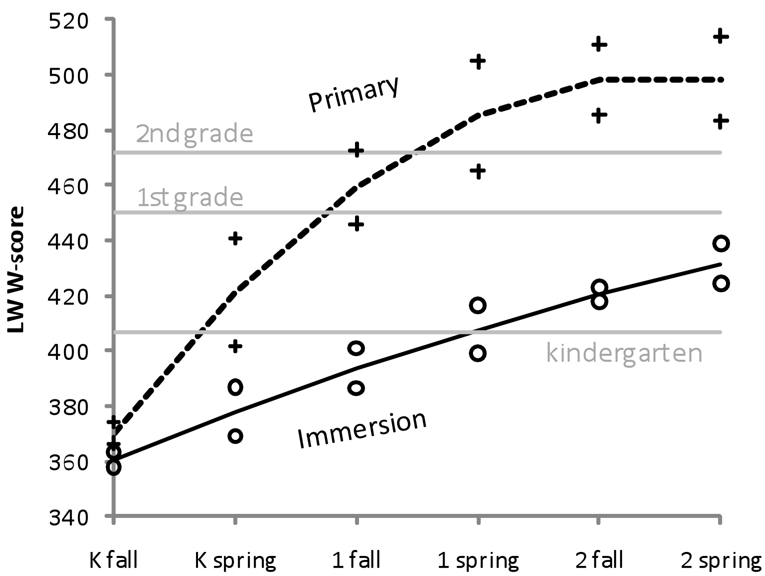



*Figure 4: Spanish model-predicted scores for each program, with ± 1 classroom SD.*




*Note*. Both panels contain estimates based on the model presented in Table 3. The markers above and below each line represent ±1 SD of classroom variability in each program. The horizontal reference lines show the end of grade expected norms for the Woodcock Language Proficiency Battery-Revised (Woodcock, 1991; Woodcock & Muñoz-Sandoval, 1995). See text for details.