

**Assessing Fidelity of Implementation of an Instructional Model:  
Targeting Student Trajectories toward Biliteracy**

Wendy Sparrow and Sandra Butvilofsky  
*BUENO Center, University of Colorado, Boulder*

Ed Wiley  
*Sequoia +Analytics, LLC*

Kathy Escamilla  
*University of Colorado, Boulder*

Please address all correspondence to:  
Wendy Sparrow  
University of Colorado, Boulder  
BUENO Center  
247 UCB, School of Education  
Boulder, CO 80309-0249  
303-735-5152  
FAX 303-492-2883  
wendy.sparrow@colorado.edu

Paper presented at the  
Annual Meeting of the American Educational Research Association  
Vancouver, Canada, April 13-17, 2012

**Assessing Fidelity of Implementation of an Instructional Model:  
Targeting Student Trajectories toward Biliteracy**

Wendy Sparrow, Sandra Butvilofsky, Ed Wiley, and Kathy Escamilla

*This study examines the degree that student outcomes associated with a bilingual/biliterate instructional model vary as a function of teachers' fidelity of implementation (FOI). It is critical to ascertain the levels at which teachers implemented the model in order to help determine its potential to foster positive trajectories toward students' bilingual/biliteracy development. We used principal components factor analysis with an oblique rotation to assess the dimensionality underlying the collection of FOI measures (i.e., the patterns of correlations between the various FOI measures), and created factor scores for each dimension represented in our final factor analytic model; these factor scores subsequently serve as the independent variables in models of student bilingual/biliteracy outcomes (in particular, scores on Spanish and English reading and writing assessments). Factor analytic results suggest eight dimensions underlying 35 FOI measures collected. Linear models of student outcomes on standard normal, regression-method scores generated for these factors suggest differential effects associated with FOI dimensions and Spanish and English reading and writing outcomes. Findings inform the revision of FOI assessment measures and improve professional development targeting program implementation. Detailed attention to FOI is critical to valid assessment of proposed educational innovations for bilingual students.*

The number of emerging bilingual students entering U.S. schools is steadily increasing (NCELA, 2010), yet bilingual program models, models empirically proven (Ramirez, Yuen, & Ramey, 1991; Thomas & Collier, 2002) to be most advantageous for such learners, are in decline. A major critique about research examining bilingual education is an absence of the examination of the quality of the instructional program (August & Shanahan, 2006; Genesee & Riches, 2006; Gersten & Baker, 2000; Slavin & Cheung, 2003). Literacy Squared<sup>®</sup> is a biliteracy/bilingual model that was specifically created to address this critique. This biliteracy project has four components: research, assessment, professional development, and an instructional component including Spanish literacy, literacy-based ESL, and cross language connections, all of which contribute to programmatic quality. While all components are important to the program writ large, this particular paper focuses on the fourth component, classroom instruction, as it is our belief that full implementation of the model at the classroom

level enhances student outcomes in reading and writing in Spanish and English, thereby ensuring students are on positive trajectories to biliteracy. In examining classroom instruction for this particular study, we analyzed teachers' lesson plans and student artifacts; conducted classroom observations; and gathered data on teachers' self-evaluations and site coordinators' and researchers' evaluations of teacher implementation. We used this information as indicators of fidelity of implementation (FOI) and investigated the relationship of these indicators to student outcomes.

The purpose of this study is to examine the degree to which student outcomes associated with a bilingual/biliterate instructional model, Literacy Squared<sup>®</sup>, vary as a function of teachers' FOI of the model. In particular, this study examines the following two questions:

1. To what degree do alternative measures of FOI represent similar or distinct aspects of teachers' implementation of Literacy Squared?
2. To what degree do students' bilingual/biliteracy outcomes vary as a function of teachers' FOI of the Literacy Squared instructional model?

### **Conceptual Framework**

The study reported in this paper was conducted to examine levels of implementation of the fourth component of the Literacy Squared model, the instructional component, in order to determine whether levels of student biliteracy achievement vary as a function of teachers' implementation of the instructional component. To this end, this study examined several different measures of FOI. The instructional component consists of three major sub-components: Spanish literacy, literacy-based ESL, and cross-language connections. Within these sub-components, quality indicators have to do with to what levels and how well teachers are implementing paired literacy instruction in Spanish and English on a daily basis, to what extent

their methods include direct/explicit instruction and include oracy, reading, writing, and metalanguage. For us, FOI exemplifies program integrity.

Examining program integrity is a critical aspect of research on educational innovations. In the context of program evaluation, FOI represents the degree to which teachers and program providers implement programs as intended by program developers (Dusenbury, Brannigan, Falco, & Hansen, 2003; Mowbray, Holter, Teague, & Bybee, 2003; Ruiz-Primo, 2006). FOI is assessed by comparing the *intended* use of a program with its *actual* use (Mills & Ragan, 2000). Given the intense ideological and practical debates about bilingual/biliterate programs, it is especially important for researchers to ensure that instructional innovations are implemented with high fidelity in order to accurately assess program efficacy. As with all educational innovations, it is critical to verify the integrity of Literacy Squared implementation to be able to appropriately attribute changes in outcomes to the program (Dane & Schneider, 1998).

Examination of program integrity requires the use of valid FOI measures. Dusenbury et. al. (2003), Ruiz-Primo (2006), and Dane and Schneider (1998) all suggest that researchers use a variety of measures to assess FOI, including: 1) Measures of adherence via self-report and/or observation protocols; 2) Measures of dosage via teacher logs or lesson plans, periodic checklists and/or other reports; 3) Quality of delivery via observation and self-report; and 4) Program differentiation with component analysis. For this study, we describe several measures developed to assess each of these four aspects of FOI of the instructional component of Literacy Squared.

This study is critical to furthering the research on the Literacy Squared program, as it is furthering our knowledge in helping us to define which critical instructional components are markers of instructional program quality as well as to ascertain the levels at which teachers need to implement these instructional components in order to help determine the model's potential in

fostering positive trajectories toward students' bilingual/biliteracy development. Findings from the study inform the revision of measures used to assess FOI, which, in turn, have implications for the improvement of professional development in the program.

### **Methods**

This study invoked quantitative methods to address the research questions presented above. In particular, we used principal components factor analysis with an oblique rotation to assess the *dimensionality* underlying the collection of FOI measures (i.e., the patterns of correlations between the various FOI measures). We created factor scores for each of the dimensions represented in our final factor analytic model; these factor scores subsequently served as the independent variables to examine the extent to which FOI impacts student bilingual/biliteracy outcomes as measured by outcome data on informal Spanish and English reading and writing assessments. This study was exploratory in nature; as such, we offer evidence that is suggestive of relationships between measures, but caution that further study is needed to confirm that these relationships are robust across study contexts.

### **Context**

This study was conducted as a part of a larger research project titled Literacy Squared. The Literacy Squared model consists of four components that collectively target trajectories toward biliteracy of emerging bilingual students in participating schools. These components include: research; assessment; professional development; and instruction, which includes Spanish literacy, literacy-based ESL, and cross language connections. Conceptually, each of these components is considered fundamental to the Literacy Squared model; moreover, each contributes to an iterative process that informs research and practice.

**Research.** Throughout the project we have collected and analyzed students' Spanish and English reading and writing data to examine student biliteracy development (Escamilla & Hopewell, 2010; Sparrow, Butvilofsky, & Escamilla, 2012) and we have conducted research on professional development and various aspects of the Literacy Squared project (Butvilofsky, 2010; Butvilofsky, & Sparrow, 2012; Soltero-González, Hopewell, & Escamilla, 2010; Sparrow, 2010).

**Assessment.** In order to establish trajectories toward biliteracy, we needed to assess students in both Spanish and English in reading and writing. This requires a holistic view of bilingualism (Grosjean, 2010; Valdés, 1991), as it is important to acknowledge what emerging bilingual children can do in both of their languages instead of viewing them in isolation from one another.

**Professional development.** Because the Literacy Squared model is unique from other bilingual models, it is essential that professional development is provided to inform teachers about the holistic theory of bilingualism that informs the instructional and assessment components. Through professional development sessions we discuss the components of the model, how to properly implement, provide demonstration lessons and on-site support, and we teach teachers how to use Spanish/English assessments to guide instruction.

**Instruction.** Instruction in Literacy Squared involves paired literacy instruction, or the teaching of both Spanish literacy and literacy-based ESL with the inclusion of cross-language connections. From kindergarten through fifth students are provided with paired literacy instruction that not only includes reading and writing, but also explicit attention to oracy and metalanguage.

### **Setting and Participants**

Literacy Squared was implemented in 16 schools in the 2010-11 academic year; these include 13 schools from one of the largest school districts in Oregon, as well as three schools from two large school districts in Colorado. This study includes 64 first through third grade teachers from these schools (50 from Oregon and 14 from Colorado). As evident in Table 1, the majority of schools have high percentages of Latino students and English language learners (ELL), as well as a large percentage of students who qualify for free or reduced priced lunch (FRL).

Table 1

School Demographic Data

School	Population	% Latino	%White	% ELL	% FRL
Austin	687	59	31	45	49
Brunswick	267	44	41	23	100
Cayden	649	85	9	75	91
Collwin	397	80	16	65	71
Forest	488	58	35	37	100
Hayden	493	85	10	61	90
Hillman	380	66	26	52	94
Holly	503	77	13	66	95
Ladow	486	70	25	46	100
Leydon	388	43	48	36	54
Marina	572	71	21	55	78
Miller	237	71	10	54	100
Morton	604	34	64	22	40
Riley	343	63	26	47	100
Sweet	570	59	28	47	100
Valley	444	90	3	72	96

Literacy Squared site coordinators serve as liaisons between the schools and the research team. Site coordinators support teachers in implementing the Literacy Squared model by modeling lessons within their classrooms and observing them as they teach. Site coordinators also assist the Literacy Squared research team in collecting data.

The Literacy Squared research team consists of seven researchers who have all been part of the project since at least 2006. Various members of the research team fulfill different roles, including: conducting the various professional development sessions with Literacy Squared teachers and site coordinators; working in classrooms; conducting classroom observations; collecting both qualitative and quantitative data; and analyzing data. Three of the four authors of this paper are part of the Literacy Squared research team.

### **Data Sources**

Four different types of program implementation data were determined to be representative indicators of level of FOI and four different assessment measures were used to examine student outcomes. Data included: 1) Lesson plan analysis; 2) Dictado analysis; 3) Site coordinator and researcher classroom observations and ratings of Literacy Squared teachers; and 4) Teacher self-evaluations. Researchers rated the four program artifacts described below to assess each participating teacher's FOI of the Literacy Squared instructional model. In each case, two raters independently assigned ratings to artifacts from a subset of 20 teachers, recalibrated in cases of inter-rater reliability of less than 80%, and repeated independent ratings. Once sufficient reliability was met for each indicator, the remaining artifacts were randomly assigned to and rated by a single rater.

**Lesson Plans.** At the beginning of each school year, teachers participating in Literacy Squared are provided dates at the beginning, middle, and end of the year by which they must submit lesson plans covering a week of instruction in their literacy blocks. These lesson plans were collected to provide evidence of the degree to which teachers understood and planned to implement the Literacy Squared model. Researchers scored lesson plans based on the various indicators within the lesson plan that teachers had to complete. The lesson plans represent an



overall view of classroom implementation of the instructional model components (Spanish literacy, literacy-based ESL, and cross-language connections) along with their sub-components and pedagogical approaches that we would expect to see in daily practice in Literacy Squared classrooms.

**Dictado Notebooks.** The Dictado was intended to be a minor method to be used in the implementation of the instructional component in Literacy Squared classrooms. This method, however, is useful and required to teach Spanish literacy, literacy-based ESL, and cross-language connections. While it is primarily a writing method, it can also be used to teach oracy, metalanguage, and reading. All Literacy Squared teachers are required to use the *Dictado* as a method to refine students' language arts skills. The Dictado teaches language arts' conventions in an integrated way, aiming to help students develop both Spanish and English language arts, and ultimately build metalinguistic skills regarding differences between the written orthographic systems of Spanish and English. It is widely used in Mexico and Central America and was adapted for use in Literacy Squared. Participating teachers are directed to implement it at least three days a week for 15-20 minutes a day, rotating weekly between Spanish and English. By the year's end, it is expected that teachers will have administered at least 15 Dictados in each language (see, e.g., Escamilla, Geisler, Hopewell, Sparrow & Butvilofsky, 2009). To collect data on Dictados, teachers were asked to submit Dictado notebooks from three randomly selected students. Researchers examined student notebooks to determine which Dictado elements were followed and rate overall Dictado implementation.

**Literacy Squared Observations.** The Literacy Squared research team observes teachers' Spanish literacy and literacy-based ESL blocks and provides written feedback to all Literacy Squared teachers once per year. Observations are rated using the *Literacy Squared Observation*

*Protocol* as a guideline to determine the nature and level of teacher implementation of the Literacy Squared model. Literacy Squared site coordinators spend a significant amount of time observing and working with teachers. Based on their knowledge of the Literacy Squared model and how teachers were implementing it, they were asked to rate teachers' program implementation in Spanish literacy and literacy-based ESL on one of three levels (high, medium, or low).

**Teacher Self-Evaluations.** All participating Literacy Squared teachers were asked to complete a form self-evaluating their levels of Literacy Squared implementation of both Spanish literacy and literacy-based ESL as high, medium, or low.

**Spanish and English Reading.** The Spanish *Evaluación del Desarrollo de Lecto-escritura* (EDL2) and the English *Developmental Reading Assessment* (DRA2) (Celebration Press, 2007a, 2007b) are parallel instruments that measure informal reading in Spanish and English, respectively. Technical properties of the EDL2 and DRA2 are described in Weber (2001) and in the DRA2 Technical Report (2009).

Scores on the EDL2 and DRA2 range from a level A through a level 60. However, increments between scores are uneven, making the calculation of a mean score inaccurate in examining student outcomes, as a mean score could be 19, and a level 19 does not exist on the assessment. Furthermore, while the primary grades have several levels within each grade, the intermediate grades do not. Thus, for the purposes of this study, scores were recalculated to better illustrate student growth and so that each whole number would represent a year of growth. A list of these adjusted scores can be found in Table 2. To remove grade-level differences in scale, each student's EDL2/DRA2 revised score was further centered on his/her grade (e.g., A first-grader's score of 1.67 corresponded to a centered score of  $1.67 - 1.00 = 0.67$ ).

Table 2

## Recalculated EDL2/DRA2 Levels

Grade Level Benchmark	Original Level	Revised Level
	A	0
	1	.25
	2	.50
End of Kindergarten	3	.75
	4	1.0
	6	1.17
	8	1.33
	10	1.5
	12	1.67
End of 1 <sup>st</sup> Grade	14	1.83
	16	2.0
	18	2.25
	20	2.5
	24	2.75
End of 2 <sup>nd</sup> Grade	28	3.0
	30	3.33
	34	3.67
End of 3 <sup>rd</sup> Grade	38	4.0
End of 4 <sup>th</sup> Grade	40	5.0
End of 5 <sup>th</sup> Grade	50	6.0
End of 6 <sup>th</sup> Grade	60	7.0

**Spanish and English Writing.** Spanish and English writing samples were collected from all students during the winter of 2011. Children were given 30 minutes to respond to a prompt provided by Literacy Squared. Spanish and English prompts were similar to facilitate comparisons of language development, but different to avoid translation. Teachers were trained to use the scoring rubric and both researchers and teachers scored these responses using a 21-point rubric covering *content*, *structural elements*, and *spelling*. (For more on the writing training and scoring, see Butvilofsky & Sparrow, 2012.)

## Results

### Factor Structure Underlying FOI Measures

Exploratory factor analysis with varimax rotation suggested eight factors; together these factors accounted for 63% of the variability in our 35 initial FOI measures.<sup>1</sup> An oblique rotation suggested similar factor structure with no bifactor correlation greater than 0.20; as such we chose to interpret the orthogonal varimax solution as optimal. Our interpretations of the eight suggested factors are detailed below.

**Factor 1: Attention to oracy, Dictado, and coherence in reading.** This factor addresses teachers' explicit attention to oracy, Dictado teaching points, and a coherent weeklong reading plan. Quality indicators for oracy indicated that objectives are apparent in both languages and are different from one another. While conceptually oracy attends to three components (dialogue, language structures, and vocabulary), the components included in this factor are Spanish and English language structures and Spanish dialogue. What is particularly noteworthy about the inclusion of oracy in this factor is teachers' overall attention to supporting oral language development, which is a major tenant of the Literacy Squared instructional component. By including teaching points for the Dictado, teachers are going beyond merely attending to spelling and punctuation, and they are addressing more specific writing and language skills. The final components included in this factor are the inclusion of a cohesive weekly reading plan in both languages and explicitly stating how the language environments will be connected.

**Factor 2: Objectives.** This factor addresses teachers' attention to creating literacy objectives that specifically address reading and writing and the inclusion of a cohesive weekly Spanish writing plan. It is important that teachers are including clear objectives that address both

---

<sup>1</sup> We considered seven- and nine-factor solutions as well, but chose to interpret the eight-factor solution as it provided the greatest degree of interpretability as well as the best balance between parsimony and explanatory power.

reading and writing, as writing has historically been a neglected aspect of literacy instruction at the elementary level. It is interesting to note that English writing was not included in this factor, though in grades 1-3, English writing is often the most neglected component, as many teachers believe that a minimum threshold in English oral language proficiency must be obtained before beginning writing instruction.

**Factor 3: Varied pedagogical approaches to teaching biliteracy.** The investigation of this factor allows us to examine if teachers are using different pedagogical approaches as they teach reading and writing in Spanish and English. When addressing quality of instruction, the Literacy Squared Theoretical Framework recommends the use of explicit/direct and interactive instructional approaches for teaching Spanish and literacy-based ESL, and the use of these approaches is different based upon the language environment and grade level. Based on this factor, teachers are incorporating approaches into their reading and writing instruction that are important for a particular grade level and language use suggested at that grade level.

**Factor 4: Teacher implementation ratings.** This factor addresses perceived levels of teacher implementation of the Literacy Squared model. In the winter, pairs from the Literacy Squared research team visit each teacher's classroom one time to observe them teaching Spanish literacy and literacy-based ESL. Based on this observation and taking into consideration the Literacy Squared Observation Protocol, the pair rates the teachers' Literacy Squared implementation level as High, Medium, or Low in Spanish literacy, as well as literacy-based ESL. A rating of Developing can also be assigned if the teacher appears to be a lower implementer, but is new to teaching and shows promise to be a higher implementer as (s)he develops as a teacher. The Literacy Squared Site Coordinators use these same ratings to rate their teachers based on their on-going work with them in and out of the classroom, and at the end of

the year, Literacy Squared teachers are asked to self-evaluate their implementation levels on this same scale.

**Factors 5-8: Dictado guidelines.** Factors 5-8 include various aspects the Dictado as the main source of evidence. Over the course of Literacy Squared implementation, we have found that the Dictado is one of the more concrete strategies for teachers to take on and implement in a more consistent way. Literacy Squared requires that teachers implement the Dictado in a systematic way, following specific guidelines that can be observed on a checklist. These four factors include the different requirements for fulfilling this cross-language strategy.

**Factor 5: Standardization and minimum requirements.** This Factor addresses if teachers adhered to the requirement to have students utilize a standard marking code for self-correcting their Dictados. It also includes whether or not teachers fulfilled the minimum number of required Spanish and English Dictados over the course of the school year.

**Factor 6: Dictado title.** Teachers are required to have a title for each Dictado.

**Factor 7: Skipping lines.** All students are required to skip lines to facilitate self-correction.

**Factor 8: Dictado consistency.** All Dictados should be administered at least three times a week.

### **Relationships Between FOI Factors and Student Outcome Measures**

Standard normal regression method factor scores for the eight factors described above served as independent variables in linear models of student Spanish and English writing and reading measures. Teachers' scores for the Dictado Guidelines factors (factors 5-8) appeared substantially bimodal in each case; as such, all models included dichotomized versions of these factors rather than their standard normal counterparts. Teacher-level outcome measures represented median scores for each teacher's students.

As this is an exploratory study, we adopted a two-step approach for each linear model. Initial models included all eight factors as independent variables. Based on results of this initial model we specified a second model that included a reduced set of variables that appeared to be significantly and/or substantially related to the modeled outcome. Results for each model are detailed below.

**Spanish Reading.** Tables 3 and 4 report parameter estimates and fit statistics for the linear model of teachers' median EDL2 scores (our measure of Spanish reading outcomes) and the FOI factors described above. In this model, four factors appeared to be related to EDL2; these four factors together accounted for  $R^2 = 30\%$  of variability in median EDL2 scores:

The strongest relationships with EDL2 scores appeared for the dichotomized Factor 5 (Dictado Guideline- Standardization and Minimum Requirements). Successfully completing minimum requirements appeared to be associated with 0.41 "years" on the rescaled and centered EDL2 (as described above). A second strong relationship with EDL2 appeared for Factor 4 (Teacher Implementation Ratings). One standard deviation in this factor appeared to be associated with 0.11 "years" on the EDL2 scale. Two other factors appeared to be *negatively* related to EDL2. These include Factor 6 (Dictado Guideline- Dictado Title) and Factor 3 (Varied Pedagogical Approaches to Teaching Biliteracy).

Table 3

Spanish Reading: Parameter Estimates from Linear Model of Teachers' Median EDL2 Scores on Fidelity of Implementation Factors

	Initial Model				Reduced Model			
	<b>b</b>	<b>SE</b>	<b>t</b>	<b>p</b>	<b>b</b>	<b>SE</b>	<b>t</b>	<b>p</b>
Intercept	.950	.342	2.779	.007				
Factor 1: Attention to Oracy, Dictado, and Coherence in Reading	-0.042	0.049	-0.853	0.398				
Factor 2: Objectives	-0.015	0.046	-0.330	0.743				
Factor 3: Varied Pedagogical Approaches to Teaching Biliteracy	-0.098	0.046	-2.138	0.037	-0.098	0.044	-2.198	0.032
Factor 4: Teacher Implementation Ratings	0.118	0.044	2.694	0.009	0.109	0.041	2.647	0.010
Factor 5 (Binary): Dictado Guideline-Standardization and Minimum Requirements	0.437	0.176	2.477	0.016	0.417	0.166	2.518	0.015
Factor 6 (Binary): Dictado Guideline-Dictado Title	-0.404	0.123	-3.278	0.002	-0.431	0.116	-3.713	0.000
Factor 7 (Binary): Dictado Guideline-Skipping Lines	-0.168	0.262	-0.642	0.523				
Factor 8 (Binary): Dictado Guideline-Dictado Consistency	0.024	0.213	0.111	0.912				



Table 4

Spanish Reading: ANOVA Table for Linear Model of Teachers' Median EDL2 Scores on Fidelity of Implementation Factors

<b>Model</b>		<b>SS</b>	<b>df</b>	<b>MS</b>	<b>F</b>	<b>p</b>	<b>R<sup>2</sup></b>
Initial	Regression	2.970	8	0.371	3.215	0.005	0.319
	Residual	6.351	55	0.115			
	Total	9.322	63				
Reduced	Regression	2.830	4	0.708	6.430	0.000	0.304
	Residual	6.492	59	0.110			
	Total	9.322	63				

### English Reading

Results for English reading were similar to those for Spanish reading. Tables 5 and 6 report parameter estimates and fit statistics for the linear model of teachers' median DRA2 scores (our measure of English reading outcomes) and our eight FOI factors. In this model, three factors appeared to be related to DRA2; these three factors together accounted for  $R^2 = 19\%$  of variability in median DRA2 scores.

The strongest relationships with DRA2 scores once again appeared for the dichotomized Factor 5 (Dictado Guideline- Standardization and Minimum Requirements). Successfully completing minimum requirements appeared to be associated with 0.72 "years" on the rescaled and centered DRA2. A second strong relationship with DRA2 also appeared once again for Factor 4 (Teacher Implementation Ratings). One standard deviation in this factor appeared to be associated with 0.15 "years" on the DRA2 scale. Finally, again similar to the EDL2 model, the dichotomized Factor 6 (Dictado Guideline- Dictado Title) appeared to be negatively related to DRA2; teachers who received "0" on this factor outperformed their counterparts by 0.46 DRA2 "years".

Table 5

English Reading: Parameter Estimates from Linear Model of Teachers' Median DRA2 Scores on Fidelity of Implementation Factors

	Initial Model				Reduced Model			
	<b>b</b>	<b>SE</b>	<b>t</b>	<b>p</b>	<b>b</b>	<b>SE</b>	<b>t</b>	<b>p</b>
Intercept	0.038	0.663	0.058	0.954				
Factor 1: Attention to Oracy, Dictado, and Coherence in Reading	-0.036	0.095	-0.379	0.706				
Factor 2: Objectives	-0.121	0.089	-1.352	0.182				
Factor 3: Varied Pedagogical Approaches to Teaching Biliteracy	-0.037	0.089	-0.414	0.680				
Factor 4: Teacher Implementation Ratings	0.148	0.085	1.746	0.086	0.129	0.080	1.619	0.111
Factor 5 (Binary): Dictado Guideline-Standardization and Minimum Requirements	0.746	0.342	2.184	0.033	0.704	0.304	2.314	0.024
Factor 6 (Binary): Dictado Guideline-Dictado Title	-0.461	0.239	-1.929	0.059	-0.492	0.217	-2.265	0.027
Factor 7 (Binary): Dictado Guideline-Skipping Lines	-0.155	0.508	-0.305	0.762				
Factor 8 (Binary): Dictado Guideline-Dictado Consistency	-0.097	0.413	-0.234	0.816				

Table 6

English Reading: ANOVA Table for Linear Model of Teachers' Median DRA2 Scores on Fidelity of Implementation Factors

<b>Model</b>		<b>SS</b>	<b>df</b>	<b>MS</b>	<b>F</b>	<b>p</b>	<b>R<sup>2</sup></b>
Initial	Regression	6.048	8	0.756	1.745	0.109	0.202
	Residual	23.834	55	0.433			
	Total	29.882	63				
Reduced	Regression	5.139	3	1.713	4.154	0.010	0.172
	Residual	24.743	60	0.412			
	Total	29.882	63				

**Spanish writing.** Models of writing outcomes evidenced different patterns of relationships than did the analogous reading models. Tables 7 and 8 report parameter estimates and fit statistics for the linear model of teachers' median Spanish writing scores and the FOI factors. In this model, three factors appeared to be related to Spanish writing; these three factors together accounted for  $R^2 = 15\%$  of variability in median Spanish writing scores.

The strongest relationship with Spanish writing scores appeared for Factor 3 (Varied Pedagogical Approaches to Teaching Biliteracy). Counter to our original expectations, that relationship is *negative*; one standard deviation in Factor 3 is associated with -0.35 points on the Spanish Writing assessment. We found some evidence suggesting relationships with Spanish writing and Factor 8 (Dictado Guideline- Dictado Consistency), and Factor 4 (Teacher Implementation Ratings). It is important to note that parameter estimates for these two factors did not meet traditional benchmarks for *statistical* significance; however, in the spirit of this exploratory study, we mention them here as candidate measures that may be informative in future studies.

Table 7

Spanish Writing: Parameter Estimates from Linear Model of Teachers' Median Spanish Writing Scores on Fidelity of Implementation Factors

	Initial Model				Reduced Model			
	<b>b</b>	<b>SE</b>	<b>t</b>	<b>p</b>	<b>b</b>	<b>SE</b>	<b>t</b>	<b>p</b>
Intercept	8.822	1.354	6.516	0.000				
Factor 1: Attention to Oracy, Dictado, and Coherence in Reading	0.039	0.194	0.201	0.841				
Factor 2: Objectives	0.222	0.183	1.217	0.229				
Factor 3: Varied Pedagogical Approaches to Teaching Biliteracy	-0.386	0.181	-2.132	0.038	-0.361	0.164	-2.202	0.032
Factor 4: Teacher Implementation Ratings	0.209	0.173	1.204	0.234	0.234	0.163	1.437	0.156
Factor 5 (Binary): Dictado Guideline-Standardization and Minimum Requirements	0.295	0.698	0.422	0.674				
Factor 6 (Binary): Dictado Guideline-Dictado Title	-0.144	0.488	-0.295	0.769				
Factor 7 (Binary): Dictado Guideline-Skipping Lines	0.398	1.038	0.383	0.703				
Factor 8 (Binary): Dictado Guideline-Dictado Consistency	1.582	0.843	1.876	0.066	1.436	0.782	1.836	0.071

Table 8

Spanish Writing: ANOVA Table for Linear Model of Teachers' Median Spanish Writing Scores on Fidelity of Implementation Factors

<b>Model</b>		<b>SS</b>	<b>df</b>	<b>MS</b>	<b>F</b>	<b>p</b>	<b>R<sup>2</sup></b>
Initial	Regression	21.704	8	2.713	1.499	0.179	0.179
	Residual	99.531	55	1.810			
	Total	121.234	63				
Reduced	Regression	18.482	3	6.161	3.597	0.019	0.152
	Residual	102.752	60	1.713			
	Total	121.234	63				

**English writing.** Table 9 and 10 report parameter estimates and fit statistics for the linear model of teachers' median English writing scores and the FOI factors. Results from these models somewhat mirror those observed for Spanish writing. In the English writing model, three factors appeared to be related to English writing; these three factors together accounted for  $R^2 = 15\%$  of variability in median English writing scores.

Similar to that of Spanish writing, the strongest relationship with English writing scores appeared for Factor 3 (Varied Pedagogical Approaches to Teaching Biliteracy). And, once again, counter to our original expectations, that relationship is *negative*; one standard deviation in Factor 3 is associated with -0.61 points on the English writing assessment. Finally, the reduced model suggested a slight relationship with Factor 5 (Dictado Guideline- Standardization and Minimum Requirements); successfully completing minimum requirements is associated with 0.41 points on the English writing assessment. With regard to this last relationship, however, we once again caution against placing too much meaning on this estimate, as it did not reach traditional benchmarks for statistical significance.

Table 9

English Writing: Parameter Estimates from Linear Model of Teachers' Median English Writing Scores on Fidelity of Implementation Factors

	Initial Model				Reduced Model			
	b	SE	t	p	b	SE	t	p
Intercept	5.823	1.919	3.034	0.004				
Factor 1: Attention to Oracy, Dictado, and Coherence in Reading	0.396	0.275	1.443	0.155				
Factor 2: Objectives	0.348	0.259	1.346	0.184				
Factor 3: Varied Pedagogical Approaches to Teaching Biliteracy	-0.611	0.257	-2.383	0.021	-0.627	0.244	-2.570	0.013
Factor 4: Teacher Implementation Ratings	-0.046	0.246	-0.185	0.854				
Factor 5 (Binary): Dictado Guideline-Standardization and Minimum Requirements	1.317	0.989	1.332	0.189	1.230	0.917	1.342	0.184
Factor 6 (Binary): Dictado Guideline-Dictado Title	-0.582	0.692	-0.842	0.403				
Factor 7 (Binary): Dictado Guideline-Skipping Lines	1.094	1.471	0.744	0.460				
Factor 8 (Binary): Dictado Guideline-Dictado Consistency	1.782	1.196	1.491	0.142				

Table 10

English Writing: ANOVA Table for Linear Model of Teachers' Median English Writing Scores on Fidelity of Implementation Factors

<b>Model</b>		<b>SS</b>	<b>df</b>	<b>MS</b>	<b>F</b>	<b>p</b>	<b>R<sup>2</sup></b>
Initial	Regression	43.556	8	5.445	1.497	0.179	0.179
	Residual	199.971	55	3.636			
	Total	243.527	63				
Reduced	Regression	25.206	2	12.603	3.521	0.036	0.104
	Residual	218.321	61	3.579			
	Total	243.527	63				

### **Limitations**

This study was purely exploratory and will help to inform future research, which will help to refine the process of examining FOI in Literacy Squared. However, it is important to note that because of the exploratory nature of this study, findings must be examined with caution, as they are not generalizable to all participants across all research sites. Furthermore, this study only included first through third grade classrooms, and further analysis is needed to determine effects of FOI on kindergarten, fourth, and fifth grade classrooms. Another possible limitation of this study is that data were not disaggregated across states, and results might be different when examining implementation independently in each state.

### **Discussion**

Literacy Squared has the potential to develop students' biliteracy skills, but in order for it to do so, it needs to be implemented with fidelity in each classroom. In considering all of the complex components of the instructional component in Literacy Squared, we hypothesized that certain factors would be more related to student outcomes than others. We thought Factors 1, 3, and 4 would have had the strongest relationship to students' reading and writing outcomes. Factors 1 and 3 were created from variables from the lesson plan artifact, which we expected would show teachers' understanding and application of them. Factor 1 contained several

variables that are unique to the Literacy Squared, including explicit attention to oracy, Dictado teaching points, and creating a connection between Spanish and English literacy instruction. However, this factor was not a significant predictor of students' reading or writing outcomes in either language.

Unlike Factor 1, which had no relationship to student outcomes, Factor 3 surprisingly had a negative relationship to Spanish reading, as well as Spanish and English writing. Because varied use of the pedagogical approaches is essential to implementation of the model, we expected this factor to be a significant predictor of student outcomes with a positive rather than a negative relationship as we promote the use of direct/explicit and interactive approaches to teaching biliteracy as opposed to teaching a skill and having students to work on it independently. Upon further inspection of this factor, we recognized two potentially problematic issues. In rating teachers on this factor, we merely looked at the number of pedagogical approaches they were using and did not consider the ways in which they were using them. For example, a teacher may have indicated the use of modeling, collaborative, and independent approaches and they would have received a score indicating their successful use of the pedagogical approaches, however we could not determine whether they were indeed using all of the approaches in their lesson, and, when using them, if they used them in a connected manner or independently of one another. Therefore, teacher ratings on this factor were not necessarily a good indicator of their implementation of the various pedagogical approaches. A second potential issue in this factor and explanation for the negative relationship is that for many teachers with higher ratings for this factor (use of more pedagogical approaches) were relying more on small group and independent approaches, rather than on explicit/direct and interactive approaches as recommended by the model. Literacy Squared does not recommend long periods



of time devoted to independent work and only recommends small group instruction minimally and in certain language environments and grade levels. Thus, the ways in which we scored teachers on the pedagogical approaches inflated teachers' ratings in an unintended way, which may have influenced the outcome; therefore we cannot conclude at this point that these approaches would, in fact, result in a negative impact on students' Spanish reading and Spanish and English writing when scored appropriately. Based on this surprising negative result, we plan to revisit this factor, how the variables within it are scored and how these variables impact student achievement.

We also expected Factor 4, Teacher Implementation Ratings, to be a significant factor for student outcomes, as two of the three variables within this factor (researcher and site coordinator ratings) were based on observations of teachers' implementation within the classroom. As expected, there was a statistically significant relationship between this factor and student outcomes, though it was associated with a relatively moderate effect of .11 years of achievement in Spanish reading and .15 years in English reading. It was not a significant predictor of writing in either language.

Factors 5-8 are all related to the Dictado, which we expected to be positively associated with students' Spanish and English writing outcomes. Factors 5, 6, and 8 all had relationships with student outcomes in reading and/or writing. The only factor that did not relate to student outcomes was Factor 7, Students Skipping Lines. Factor 5, Standardization and Minimum Requirements, was the strongest indicator of Spanish and English reading, and it had a slight relationship with English writing. That this had such a strong predictor of reading was unanticipated, as the Dictado is a strategy to improve student writing. Nonetheless, this finding reinforces our understandings of the reciprocal relationship between reading and writing (Clay,

1991; Ferreiro, 1986). Factor 8, Dictado Consistency, was associated with an increase of 1.41 points in Spanish writing. While this was not surprising, we would have also expected a similar impact in English writing. However, it was not a significant predictor of English writing, which may be due to the fact that not enough of our first through third grade teachers implement the Dictado in English with consistency so that such a relationship would not appear in English writing. The final Dictado factor that was a significant predictor of Spanish and English reading outcomes was Factor 6, Dictado Title. While this relationship was negative, which was unexpected, upon further inspection of the data, we determined that many of the teachers that did not use a title in the Dictado were first grade teachers. From conversations with some of these first grade teachers, we know they were reluctant to use a title as they felt the Dictado was so short it did not necessitate one. While we know that the Dictado should have a title, we do not believe this finding is indicative that not using a title on the Dictado will provide higher student reading outcomes.

### **Significance of the Study**

This study contributes generally to research on how to assess FOI of educational innovations. In the broad field of educational research, researchers involved examining FOI measurement have concluded that at this point, the field is more theoretical than practical (Bond et. al., 2000; Dane & Schneider, 1998; Dusenbury et. al., 2003; Mowbray et al., 2003; Ruiz-Primo, 2006). These same researchers conclude that even when researchers report the steps taken to *promote* FOI, they often fail to describe methods used to *assess* it. Moreover, FOI assessment is rather recent in K-12 curriculum intervention research even though its use in program evaluation dates back 35 years (O'Donnell, 2008). With regard to this study, we attempted to add to the rather scant literature on how to assess FOI in K-12 curriculum intervention. In this case,

our analysis led us to conclude that the implementation of a direct and explicit method in the form of the Dictado, when fully implemented, had a very positive impact on reading outcomes in both Spanish and English. This finding is significant in that it provides evidence at the qualitative level of implementation with regard to dosage and not simple presence or absence of a strategy or method.

Furthermore, our findings indicate that observations of classroom instruction and the subsequent assigning of levels of implementation, in this case, high, medium, and low, are reliable in predicting student outcomes. This is significant for the overall project in terms of the need for professional development to prepare site-based observers and future Literacy Squared researchers to observe and rate instruction. Examining how observations of teachers' instruction related to levels of implementation and student outcomes also contributes to the body of research on FOI.

This study contributes in particular to the field of bilingual/dual language education particularly in the area of literacy. Findings from this study, especially with regard to the results of the Dictado, its level of implementation, and student outcomes adds to research conducted by Genesee and Riches (2006) who suggested that direct and explicit methods of instruction are beneficial for literacy outcomes in emerging bilingual learners. The Dictado is a direct and explicit approach to teaching. A second and very important finding is the impact that full implementation of the Dictado appears to have on reading outcomes. While primarily a method for teaching students to write, the Dictado appears to have benefits in reading outcomes as well. Given the narrow focus on teaching reading in many U.S. schools including bilingual/dual language programs, these findings are significant and add to the research indicating that we can teach reading by teaching writing, and perhaps, in programs such as Literacy Squared, we need

to continue to emphasize the utilization of methods such as the Dictado as a means of enhancing reading as well as metalinguistic awareness. The consensus of syntheses of research in the field of bilingual/dual language education have found that it is often not possible to draw conclusions about the overall efficacy of programs because of the lack of information about FOI reported by researchers (August & Shanahan, 2006; Slavin & Cheung, 2005). Detailed attention to FOI issues is critical to the valid assessment of the effectiveness of proposed educational innovations for emerging bilingual students. We feel that this study has significance not only because of its attention to FOI and how to assess it, but also because the outcome measures include both Spanish and English reading and writing. However, we caution that this was an exploratory study and more research is needed on FOI and its effects on program efficacy.

## References

- August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the national literacy panel on language-minority children and youth*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Bond, G. R., Evans, L., Salyers, M. P., Williams, J., & Kim, H. (2000). Measurement of fidelity in psychiatric Rehabilitation. *Mental Health Services Research, 2*, 75-87.
- Butvilofsky, S. A. (2010). *Towards the development of a biliterate pedagogy: A case study of emerging biliterate writing*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses. (UMI No. 3419442).
- Butvilofsky, S., & Sparrow, W. (2012). Training teachers to evaluate emerging bilingual students' biliterate writing. *Language and Education, 1*, 1-21.
- Celebration Press. (2007a). *Evaluación del desarrollo de la lectura*. Parsipanny, NJ: Celebration Press.
- Celebration Press. (2007b). *Developmental Reading Assessment*. Parsipanny, NJ: Celebration Press.
- Clay, M. M. (1991). *Becoming literate: The construction of inner control*. Portsmouth, NH: Heinemann.
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*, 23-45.
- Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18*, 237-256.

- Escamilla, K., & Hopewell, S. (2010). Transitions to biliteracy: Creating positive academic trajectories for emerging bilinguals in the United States. In J. Petrovic (Ed.), *International perspectives on bilingual education: Policy, practice, and controversy* (pgs. 69-93). Charlotte, NC: Information Age Publishing Series.
- Escamilla, K., Geisler, D., Hopewell, S., Sparrow, W., & Butvilofsky, S. (2009). Using writing to make cross-language connections from Spanish to English. In C. Rodriguez (Ed.), *Achieving literacy success with English language learners*, (pgs. 141-156). Columbus, OH: Reading Recovery Council of North America.
- Ferreiro, E. (1986). The interplay between information and assimilation in beginning literacy. In W.H. Teale and E. Sulzby (Eds.), *Emergent literacy: Writing and reading*, (pgs.15-49). Norwood NJ: Ablex.
- Genesee, F., & Riches, C. (2006). Literacy: Instructional issues. In F. Genesee, K. Lindholm-Leary, W. Saunders, and D. Christian (Eds.), *Educating English language learners: A synthesis of research evidence*. Cambridge: Cambridge University Press.
- Gersten, R., & Baker, S. (2000). What we know about effective instructional practices for English-language learners. *Exceptional Children*, 66, 454-470.
- Grosjean, F. (2010). *Studying bilinguals*. New York: Oxford University Press.
- Mills, S. C., & Ragan, T. J. (2000). A tool for analyzing implementation fidelity of an integrated learning system. *Educational Technology Research and Development*, 48(4), 21-24.
- Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, 24, 315-340.
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation

and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, 78, 33-84.

Pearson Education, Inc. (2009). *DRA2 technical manual: Developmental reading assessment*. Upper Saddle River, New Jersey: Pearson Education, Inc. Retrieved: [http://s7ondemand7.scene7.com/s7ondemand/brochure/flash\\_brochure.jsp?company=PearsonEducation&sku=DRA2\\_TechMan&vc=instanceName=Pearson&config=DRA2\\_TechMan&zoomwidth=975&zoomheight=750](http://s7ondemand7.scene7.com/s7ondemand/brochure/flash_brochure.jsp?company=PearsonEducation&sku=DRA2_TechMan&vc=instanceName=Pearson&config=DRA2_TechMan&zoomwidth=975&zoomheight=750)

Ramírez, D., Pasta, D. Yuen, S., & Ramey, D. (1991). *Final report. Longitudinal study of structured English immersion strategy, early-exit and late-exit transitional bilingual education programs for language-minority children*. (Vols. 1 & 2). San Mateo: CA: Aguirre International.

Ruiz-Primo, M. A. (2006, February). *A multi-method and multi-source approach for studying fidelity of implementation* (CSE Report 677). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Slavin, R. & Cheung, A. (2003). *Effective reading programs for English language learners: A best-evidence synthesis*. Washington DC: Center for Research on the Education of Students Placed at Risk (CRESPAR).

Soltero-González, L., Hopewell, S., & Escamilla, K. (2010). A bilingual perspective on writing assessment: Implications for teachers of emerging bilingual writers. In G. Li and P. A. Edwards (Eds.), *Best practices in ELL instruction*, (pgs. 222-244). New York: The Guilford Press.

- Sparrow, W. (2010). *An examination of the challenges associated with fidelity of implementation in an innovative biliteracy program*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses. (UMI No. 3408973).
- Sparrow, W., Butvilofsky, S., & Escamilla, K. (2011). The evolution of biliterate writing through simultaneous bilingual literacy instruction. In E. Bauer & M. Gort (Eds.), *Early biliteracy development: Exploring young learners' use of their linguistic resource*, (pgs. 157-182). New York: Routledge.
- Thomas, W.P., & Collier, V.P. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement*. Santa Cruz, CA, and Washington, DC: Center for Research on Education, Diversity & Excellence.
- Valdés, G. (1991). Bilingual minorities and language issues in writing: Toward professionwide responses to a new challenge. *Written Communication*, 9, 85-136.
- Weber, W. (2001). *Developmental Reading Assessment and Evaluación del desarrollo de la lectura: A Validation Study*. Retrieved: [www.pearsonlearning.com](http://www.pearsonlearning.com).