Spelling Development in Spanish–English Emergent Bilinguals Between Grade 1 and Grade 3: Exploring Scoring Methods and Instructional Contexts

Introduction

For emergent bilingual children, learning spelling in two different languages imposes a greater demand than monolingual children (Koda, 2007). The two languages interact with each other, necessitating children having to attune to the language properties that are specific to each language. For Spanish-English bilinguals, their L1 (Spanish) has many similarities with English but has a relatively shallow orthography compared to English which has important implications for learning how to spell (Dixon et al., 2010; Katz & Frost, 1992; Pérez Cañado, 2005). Extant research on Spanish-English bilingual children's spelling errors indicate that there are bidirectional crosslinguistic influences (i.e., transfer of linguistic or orthographic features from one language to another; Koda, 2005) in both directions (Bahr et al., 2015; Wolters & Kim, 2023; Zaretsky, 2020). Then, the next crucial step is to understand whether their overall spelling development in Spanish and English support or facilitate development in the other, and if so, in which direction.

When examining spelling development, one aspect to consider is how to score spelling. Recent studies have demonstrated that using a more fine-grained method for scoring spelling can reveal a more nuanced understanding of spelling development (Clemens et al., 2014; Lee & Al Otaiba, 2017; Treiman et al., 2019), yet most studies to date have focused exclusively on English, with limited attention to other languages. Another aspect to consider is the language of instruction used in bilingual children's instructional program (e.g., English immersion or English-Spanish dual immersion). According to the input driven theory, the amount of exposure to each language is an important factor that influences children's spelling development (Ellis, 2002). To this end, the current study aimed to investigate how bilingual children's spelling develops from Grade 1 to Grade 3 and how Spanish and English spelling are interrelated over time. We also examined whether the relation varies as a function of which scoring method is used for spelling and which instructional program children attend.

Theoretical Frameworks on Bilingual Spelling Development

Spelling refers to accurate production of words in written form or the process of transcribing words into their written form by using the appropriate sequence of letters (Ehri, 2000). Spelling development is challenging for children as it requires the combination of precise knowledge of linguistic skills (Ehri, 1987; Moats, 2009; Treiman, 2017). Theoretical models of spelling (Treiman, 1993) emphasize that several foundational skills are essential for spelling: phonological awareness (awareness of words' phonological structure), orthographic awareness (understanding of orthographic symbols and permissible letter patterns), and morphological awareness (knowledge of meaningful word parts). In fact, extensive research has demonstrated the importance of these foundational skills in learning how to spell in English (Apel et al., 2012; Caravolas et al., 2005; Deacon et al., 2009; Kim et al., 2013; Levesque et al., 2021).

For bilingual children, this complex process of spelling development increases due to the interaction between two different languages, necessitating adjustments to language specific demands (Koda, 2007). One important concept to understanding literacy development of bilingual children is the distinction between language-general and language-specific skills. Koda (2007) emphasizes the significance of identifying these skills to comprehend the complexities of language development, particularly in the context of biliteracy. Important language-general skills underlying spelling include metalinguistic awareness such as phonological, orthographic, and morphological awareness. Once a child develops phonological awareness in one language, it can be used to learn to spell in the second language. Orthographic awareness and morphological awareness are also language-general skills in that one's understanding and the ability to recognize and manipulate morphemes can transfer across languages, even though specific orthographic or morphological features may differ.

Studies on Spanish-English bilinguals have revealed that these language-general skills are indeed transferrable and facilitate the development of a second language (phonological awareness: Branum-Martin et al., 2006; Vaughn et al., 2006; Wawire & Kim, 2018; orthographic awareness: Deacon et al., 2013; morphological awareness: Kuo et al., 2017). However, there are also language-specific traits that demand bilinguals' sensitivity to both languages. For instance, a language-specific skill for Spanish-English bilingual children would be attuning to the optimal grain size for decoding or spelling according to the Psycholinguistic Grain Size theory (Ziegler & Goswami, 2005). The required grain size for Spanish is smaller at the phonemic level than English since Spanish has a consistent phoneme-grapheme correspondence (Katz & Frost, 1992).

The presence of language-general and -specific skills give rise to both positive and negative transfer in biliteracy development (Figueredo, 2006). Cross-linguistic transfer refers to the "automatic activation of well-established L1 competencies (mapping patterns) triggered by L2 input" (Koda, 2007, p. 17). According to the Linguistic Interdependence Theory (Cummins, 1979, 1992), one's L1 can support learning of L2 through the transfer of shared features between the two languages. The developmental interdependence hypothesis further posits that the level of proficiency in L2 is influenced, in part, by the proficiency in L1. Cummins (1979) stresses the significance of reaching a certain threshold of proficiency in both languages to avoid limited or negative transfer. On the other hand, the Unified Model of Bilingualism (MacWhinney, 2005) assumes that the less developed language is dependent on the dominant language, suggesting a one-way influence from L1 to L2. For instance, children whose L1 is an alphabetic language may apply phoneme-grapheme correspondences from their L1 to English (Dixon et al., 2010). These frameworks illuminate the intricate interplay of language dynamics in biliteracy development.

Spelling Development of Spanish-English Bilinguals

Spanish and English are two languages that share many similarities, although important differences do exist (Pérez Cañado, 2005). Understanding the characteristics of their respective writing systems can provide insights into the spelling development of Spanish-English bilinguals. Both languages follow the alphabetic system, and the majority of letters in the English alphabet are also present in the Spanish alphabet, with similar sounds. Additionally, they share cognates, which are words that possess phonological, orthographic, and semantic similarities across languages (Rubin & Carlan, 2005; Wolters & Kim, 2023). For example, *familia* in Spanish and *family* in English are considered cognates. Bilingual children may find it easier to learn how to spell new English words that are cognates with Spanish words, as they share similarities in form, pronunciation, and meaning which can lead to positive transfer.

An important difference between Spanish and English is the variation in orthographic depth (i.e., the consistency of grapheme–phoneme correspondences in a language; Gaintza & Goikoetxea, 2016; Katz & Frost, 1992). The Orthographic Depth Hypothesis (Katz & Frost, 1992; Seymour et al., 2003) posits that the difficulty of learning a language depends on the consistency of a language's orthographic system (Georgiou et al., 2012). Spanish has a relatively simple grapheme-to-phoneme correspondence with 29 graphemes and 24-25 phonemes (Gaintza & Goikoetxea, 2016). While there are some inconsistencies where some phonemes can be represented by more than one grapheme without following any particular rules, most of the words follow regular spelling rules. In contrast, English is considered a deep orthography that has many inconsistencies in its phoneme-grapheme mapping. For instance, in Spanish, the phoneme /f/ is consistently represented by the letter *f* (e.g., *farmacia*) whereas English represents this sound with both *f* and *ph* (e.g., *fish* vs. *pharmacy*). This difference in orthographic depth makes learning to spell in Spanish easier than in English, and bilingual children whose first language is Spanish may face more challenges in learning English spelling (Figueredo, 2006). They might initially try to apply phoneme-grapheme correspondences they know from Spanish until they gradually learn that English has more irregularities (Cummins, 1979). Other differences include Spanish letters such as *ñ* (eñe) which do not exist in English, and the use of accent marks on vowels (e.g., á) in Spanish which are also not used in English spelling. Furthermore, the pronunciation of the letters when reading words may also differ. For instance, the word *hospital* has the same spelling in English and Spanish but the pronunciation in English is /hɒs.pɪ.təl/ whereas *h* is silent in Spanish (/ospiˈtal/).

Several studies examining Spanish-English bilingual children's spelling errors have indeed found cross-linguistic relations consistently; from Spanish to English (Bahr et al., 2015; Fashola et al., 1996; Hevia-Tuero et al., 2023; Lindner et al., 2022; Rubin & Carlan, 2005; Wolters & Kim, 2023; Zutell & Allen, 1988) and also from English to Spanish although smaller in volume (Bahr et al., 2015; Wolters & Kim, 2023). In addition, some studies have looked at the characteristics of bilingual children's overall spelling development. A review of studies examining the influence of L1 on English (L2) by Figueredo (2006) highlighted the opportunity for transfer of underlying linguistic processes between L1 and L2 in alphabetic languages (Cummins, 1979; Deacon et al., 2013). In line with this review, Rubin and Carlan (2005) analyzed the writing in Spanish and English of bilingual children aged 3 to 10 and found overall developmental stages in both languages to be similar, although one language might be at a later stage than the other. In fact, a study on Spanish monolingual children found that their process of Spanish spelling development is similar to that in more opaque languages like English, suggesting common underlying mechanisms for spelling development in alphabetic languages with varying orthographic transparency (Defior & Serrano, 2005). The spelling development of bilingual learners do not seem to differ significantly from monolingual learners, indicating a common developmental pathway.

Overall, theoretical frameworks and previous studies suggest that transfer between L1 and L2 occurs in learning to spell in multiple languages. Despite this unique aspect of bilingual children, most research on spelling development focuses on monolingual English-speaking children, with limited studies on Spanish-English bilinguals. Existing research that has been done with this population primarily focused on categorizing individual spelling errors to understand cross-linguistic relations (Bahr et al., 2015; Fashola et al., 1996; Hevia-Tuero et al., 2023; Lindner et al., 2022; Rubin & Carlan, 2005; Wolters & Kim, 2023; Zutell & Allen, 1988). Although these studies provide valuable insights, this approach leaves a gap in understanding whether spelling proficiency in one language will support or impede the other over time. Building on prior research that has primarily examined bilingual learners' spelling errors at a micro-level, there is a need for studies examining the relation between the broader Spanish and English spelling proficiency, especially during the early grades when literacy instruction increases. Such research can clarify whether spelling development in one language supports growth in the other, even in the presence of cross-linguistic spelling errors.

Using a Nonbinary Scoring Method for Bilingual Children's Spelling

According to the dynamic relations hypothesis of Direct and Indirect Effects Model of Writing (DIEW; Kim, 2020; Kim & Graham, 2022), how component skills (e.g., spelling, oral language, word reading) relate to writing may vary as a function of how constructs are measured. One example is written composition. For instance, second grade students' written composition was evaluated on the quality, productivity (text length), and correctness, spelling was more strongly related to correctness in compositions while handwriting fluency was more strongly related to productivity (Kim & Graham, 2022). Similarly, previous research has shown that depending on how children's spelling is scored, the relation between English spelling and reading skills varied in directionality or magnitude (Caravolas et al., 2001; Clemens et al., 2014). For instance, Caravolas and colleagues (2001) demonstrated that the correlation between early spelling and later reading was positive and strong when phonological spelling scores (phonological plausibility of the spelling ranging from 0 to 4) were used, whereas the relation was weaker or sometimes nonsignificant when correctness scores were used. In Clemens et al. (2014), nonbinary spelling scores in kindergarten that gave partial credit to imperfect spellings explained 6% to 15% more variance of children's reading in Grade 1 compared to correctness scores. Extending this reasoning to bilingual populations, different relations between Spanish and English spelling may emerge when more than one scoring method for spelling is used.

Scoring children's spelling dichotomously as either correct or incorrect (i.e., correctness scores) overlooks the nuances of their developing knowledge (Clemens et al., 2014; Lee & Al Otaiba, 2017). In addition, correctness scores can lead to floor effects, especially with younger children who often spell words partially correct (Caravolas et al., 2001; Ehri, 1987). This applies across monolingual and multilingual children, including Spanish-English emergent bilingual children (Dixon et al., 2010; Verhoeven, 2000). Correctness scores alone may not fully capture their spelling development as it does not account for different degrees of correctness present in their attempted spelling. Using a nonbinary scoring method for bilingual children's spelling can provide a more nuanced understanding of their spelling proficiency because it does not disregard partially correct spellings (Lee & Al Otaiba, 2017; Ritchey et al., 2010).

One method of nonbinary scoring for spelling is *text distance* which measures the degree of incorrectness of an attempted spelling. These scores are calculated using computerized algorithms that indicate the distance between two words. Thus, higher text distance scores indicate more distance from the correct spelling, or a more incorrect spelling. Because this method is fully automated, it avoids time- and labor-intensive manual scoring compared to other human-scored nonbinary scores. Another advantage of text distance is that various algorithms can be used, each emphasizing different aspects of spelling (e.g., phonological or orthographic distance). For example, phonology-based distance scores are solely based on the pronunciation of the strings, so a misspelling like *spel* for *spell* would receive full credit for phonological plausibility as each phoneme is correctly represented despite the error. In contrast, orthography-based distance measures would penalize *spel* for omitting the letter *l*. In this sense, correctness and text distance scores both measure children's spelling ability but capture different aspects. Conventional correctness scores categorically (correct or incorrect) reflect whether a child can accurately reproduce the spelling of an entire word. In contrast, text distance scores capture partial developing spelling knowledge by measuring how close an attempted spelling is to the target word. These scores also indirectly reflect foundational skills involved in spelling development—such as phonological awareness and orthographic awareness (Treiman, 1993). Accordingly, text distance provides a continuous, fine-grained view of children's spelling development that can potentially complement what is captured by correctness scores.

Although text distance scoring algorithms have been developed based on English, orthography-based algorithms can be effectively applied to alphabetic languages like Spanish since the scoring is conducted at the letter level. On the other hand, the pronunciation differences between English and Spanish (e.g., the letter /h/ is silent in Spanish) makes phonology-based algorithms less suitable for Spanish. Thus, in the current study, we use two different types of text distance scores that are orthography-based (e.g., Damerau-Levenshtein and cosine distance; see Methods section for more detail on these scores). Only a handful of studies have used text distance scores for examining English spelling (Authors, 2024; Treiman et al., 2016, 2019), and no research to our knowledge has yet investigated the use of text distance scores for the spelling of bilingual populations where both languages are alphabetic. Because cross-language transfer often occurs at the level of underlying foundational literacy skills (e.g., phonological and orthographic awareness), a more sensitive, nuanced scoring method may offer a comprehensive understanding of Spanish-English bilingual children's spelling development and uncover different relations between Spanish and English spelling.

Influence of Instructional Programs

According to the input-driven theory (Ellis, 2002), language learning is closely aligned with the quantity and quality of linguistic input learners receive. One of the most critical sources of such input for literacy development is the language of instruction children receive (Koda, 2007). In bilingual education contexts, instructional language not only provides linguistic exposure but also determines what features of language are explicitly taught and practiced in academic settings. If a bilingual child attends an English immersion program, we can expect them to develop greater progress in English spelling than those in English-Spanish dual immersion programs where part or entire instruction focuses on Spanish (language of instruction and target literacy skills are in Spanish). In fact, previous research on Spanish-English bilingual children supports the moderating effect of instructional programs on linguistic relations (Kuo et al., 2017; San Francisco et al., 2006; Wolters & Kim, 2023). In a study on Grade 1 Spanish-English bilingual children, San Francisco and colleagues (2006) found Spanish-influenced spelling only in children receiving Spanish instruction whereas those who received English instruction did not. Extending this finding, Wolters and Kim (2023) found that Spanish-English bilinguals exhibited English-influenced spelling errors if they attended English immersion programs. Therefore, the language of instruction children receive at instructional programs is an important factor to consider, as it is anticipated to impact spelling growth and the interrelation between L1 and L2 for bilingual children.

Present Study

The primary aim of the current study is to enhance our understanding of the spelling development of Spanish-English emergent bilingual children by examining the relation between their spelling in both languages and how they interrelate over time. While previous research on Spanish-English bilingual children's spelling error analysis has revealed frequent cross-linguistic transfer, no studies to our knowledge have examined how spelling in both languages in earlier time point contributes to later spelling skill in the other. In addition, this study is the first attempt to explore the use of a nonbinary scoring measure (e.g., text distance) for assessing Spanish spelling. As prior studies have shown advantages of using nonbinary scores for children's spelling in English, it would be insightful to see whether this method is appropriate for examining Spanish spelling as well. Additionally, the study examined whether language of instruction (i.e., English immersion vs. dual immersion) affects spelling development and the interrelation between the two languages. The present study draws on multiple frameworks to examine different aspects of bilingual spelling development (e.g., cross-linguistic transfer, measurement of spelling [different scoring methods], and the role of instructional programs). These frameworks collectively guide our interpretation of the developmental relations observed between Spanish and English spelling. The research questions that will be addressed are as follows:

1. How do spelling skills in English and Spanish of Spanish-English bilingual children change from Grade 1 to Grade 3?
    a. Do their spelling skills differ by instructional context?
2. How are Spanish and English spelling interrelated over time from Grade 1 to 3?
3. Does the relation between Spanish and English spelling differ based on which spelling scoring method (e.g., correctness scores vs. text distance scores) is used?
4. Does the relation between Spanish and English spelling differ based on the instructional program in which children are enrolled?

Hypotheses

For research question 1, we hypothesized that children's spelling skills in both English and Spanish would improve from Grade 1 to Grade 3. We expected minimal differences in children’s English and Spanish spelling in Grade 1 depending on which program they have been attending, but relatively large differences in Grade 3. Specifically, we expected children participating in an English-Spanish dual immersion program would show greater improvements in Spanish spelling by Grade 3 compared to their peers in an English immersion program and vice versa. Regarding the interrelation between Spanish and English spelling (RQ2), we hypothesized a significant bidirectional relation between the two languages over time given prior studies showing cross-linguistic transfer in both directions. Additionally, we expected that compared to binary correctness scores, the use of a more detailed scoring method (i.e., text distance capturing the degree of correctness) would reveal a different pattern of interrelation (RQ 3), although we do not have specific predictions on the nature of these changes. Lastly, when examining whether these relations vary by instructional program, we expected a stronger relation from Grade 1 Spanish to Grade 3 English for children in dual immersion programs and a stronger relation from Grade 1 English to Grade 3 Spanish for children in English immersion programs (RQ 4).

Methods

Participants

The current study includes 209 Spanish-English bilingual children (42.7% male) that participated in a larger study (Author, 2025). Only children who provided assent were included in the study, in accordance with approval from the Institutional Review Board (Institution masked for blind review). The children were from a high poverty school district in the Southwestern region in the U.S. The majority of children were of Hispanic descent (81.5%) and most of the children (80.1%) were designated as English Learners (ELs) by the school district (see Table A1 in the Appendix for the number of ELs by instructional program). About 73% of the children were eligible for the free and reduced-price lunch program. Nineteen children were receiving disability services from the school and the majority were for language speech impairment (n = 11).

Children attended either an English immersion program or an English-Spanish dual immersion program (hereafter referred to as dual immersion program). About one-third of the children (38.4%) were in an English immersion program while the remaining children (61.6%) were in a dual immersion program. Five children changed from a dual immersion program in Grade 1 to an English immersion program in Grade 3. Among children in an English immersion program, 54% were ELs and 68% were ELs in dual immersion programs. English learner status information was obtained from district records and was determined based on students' performance on state English proficiency tests for ELs, as well as input from teachers and parents. The English immersion program used the Benchmarks Events curriculum, and all instruction was given in English. In the dual immersion program, the same curriculum was used in English and Spanish versions where 90% of the instruction was given in Spanish in Grade 1 and 80% in Spanish for Grade 3.

Spelling Assessment

The children were assessed for their spelling in Spring once in Grade 1 and once more in Grade 3. The numbers of students for each spelling test by instructional program are given in the Appendix (see Table A2). For the English spelling test, children were given 18 developmentally appropriate words based on prior work (Masked for blind review). For the Spanish spelling test, 15 words were adapted from the Palabras a su Paso (Helman et al., 2013). In both tasks, the target word was first presented orally in isolation, once more in a sentence, and presented in isolation one last time. For both languages, the same word lists that were used to assess children in Grade 1 were used again two years later in Grade 3. The complete list of spelling words for English and Spanish is given in the Appendix (see Table A3). The Cronbach's α of the English spelling tests were 0.91 and 0.93 for Grades 1 and 3, respectively. For the Spanish spelling tests, the Cronbach's α were 0.79 and 0.84 for Grades 1 and 3, respectively.

Spelling Scoring Procedure

All four spelling tests (i.e., Grade 1 English and Spanish, Grade 3 English and Spanish) were transcribed by three trained research assistants that were Spanish-English bilingual undergraduate students majoring in education. The transcribers typed the spelling entries verbatim, and all the spelling tests were double transcribed in pairs to ensure accuracy since illegible handwriting by younger bilingual students was frequently observed. Once two research assistants transcribed each spelling test item, the interrater reliability was calculated and the disagreements for transcription was resolved after discussion. The interrater reliability was 92% for Spanish spelling tests and 93% for English spelling tests.

Children's correctness and text distance spelling scores were obtained using a researcher-developed website (Masked for blind review; Authors, 2023). This website provides a user-friendly tool for scoring transcribed spelling both dichotomously and for text distance scores by using the Python library 'textdistance'. Among the various text distance algorithms available, only letter-based algorithms were chosen since the sound-based algorithms are not applicable to Spanish. Two algorithms, Damerau-Levenshtein and cosine distance, were chosen based on previous studies using text distance for children's spelling and matching misspelled strings (Authors, 2024; Moreau et al., 2008; Treiman et al., 2019).

Damerau-Levenshtein determines the minimum number of edits (addition, deletion, substitution, transposition) needed to convert an incorrect spelling into the correct one (Kruskal, 1983 as cited in Treiman et al., 2016). For example, for the target word *bird*, attempted spelling including an insertion (*birld*), a deletion (*bid*), a substitution (*berd*), and a transposition (*brid*) would get a Damerau-Levenshtein score of 1 (i.e., one edit required to make each spelling correct). To allow comparison across words of different lengths, normalized Damerau-Levenshtein scores ranging from 0 to 1 were obtained by dividing the number of edits by the maximum possible distance, or the target word's length. Next, cosine distance evaluates the frequency of each letter in a spelling without considering the order of the letters (i.e., order-insensitive measure capturing the presence of correct letters). Cosine distance converts each word into a vector based on its letter frequency and calculates how far apart these vectors are. For instance, the misspelling *gril* for the word *girl* would get a cosine distance score of 0, indicating no distance since all letters are present but out of order. The cosine distance scores range from 0 to 1 and did not require normalization. Damerau–Levenshtein, in contrast, penalizes such positional errors. Using a composite score that averages Damerau–Levenshtein and cosine distance helps avoid over-penalizing minor spelling errors common among early bilingual learners. Cosine distance is insensitive to letter order compared to edit-based measures like Damerau–Levenshtein, making it particularly useful for capturing developmental spellings in which letters are rearranged (e.g., gril for girl). By combining both metrics, the composite score accounts for specific letter-level inaccuracies as well as overall similarity in word composition, providing a balanced and developmentally appropriate measure of spelling performance.

Given this is the first study to attempt using text distance algorithms on Spanish spelling, test rounds with Spanish words were conducted to examine whether the algorithms processed Spanish-specific letters correctly. These tests involved inputting Spanish words that contained letters not present in English (e.g., ñ) to verify that the software recognized them as distinct letters. The test analyses confirmed that letters unique to Spanish were processed equivalently to English letters. For instance, the Damerau–Levenshtein distance between the misspelling *extrano* and the target word *extraño* was 1, reflecting a single substitution (n → ñ). Cosine distance, which scores spelling irrespective of order, treated *extrano* and *extramo* as equally distant from *extraño*, since both differ from the target by one letter. Once both text distance scores were obtained, the average scores for Spanish and English spelling tests for each grade as a composite text distance score were calculated. Example of children's spellings and their respective text distance scores are provided in the Appendix (see Table A4).

Data Analytic Strategies

Children's spelling data were first examined descriptively to evaluate the normality of both correctness and text-distance scores and to evaluate changes in English and Spanish spelling skills among Spanish–English bilingual children from Grade 1 to Grade 3 (RQ 1). The correlations between English and Spanish spelling scores were also examined. We then used cross-lagged panel models (CLPMs) to estimate directional associations between English and Spanish spelling across time while controlling for autoregressive stability. It is important to note that the current study is examining change over two time points within a longitudinal perspective. Although random-intercept CLPMs (RI-CLPMs; Hamaker et al., 2015) better separate within- and between-person variance, they require at least three measurement occasions. We also considered latent change score (LCS) models, which can be estimated with two waves but do not allow simultaneous estimation of autoregressive relations, one of the central focuses of the present study. With two time points, the standard CLPM provides interpretable estimates of stability-adjusted relations, which we interpret as directional associations rather than causal within-person effects (Lucas, 2023).

A CLPM was fit using correctness scores where Grade 1 English and Spanish spelling predict Grade 3 English and Spanish spelling to examine how Spanish and English spelling is interrelated over time (RQ 2).

Another CLPM was fit using the composite of text distance scores to see whether the relation between Spanish and English spelling differs based on which spelling scoring method is used (RQ 3). Finally, multi-group analyses were conducted using both correctness and text distance scores to see whether the relation between Spanish and English spelling differs based on which instructional program children are in (RQ 4). For the multigroup analyses, all pathways were initially constrained to be equal across instructional programs to test for model invariance (whether the relations among variables differed significantly between instructional contexts). If the fully constrained model showed evidence of misfit, pathways were then sequentially freed to examine whether allowing specific pathways to vary significantly improved model fit (indicating difference between instructional programs).

As a sensitivity check, we conducted additional analyses where an identical set of models were fit that (1) clustered standard errors at the school level to account for the nesting of students and (2) excluded the five children who changed instructional programs between Grades 1 and 3. The substantive pattern of results remained unchanged and the simpler models were reported for clarity and ease of interpretation. Stata Version 16 was used for running descriptive statistics and correlation analysis and Mplus Version 8 was used for running CLPMs. All Mplus syntax and output files associated with the CLPM analyses—including model specifications, estimation results, and sensitivity checks—are publicly available at: *Github link masked for blind review*.

Approximately 10% of the children were missing their English and Spanish spelling tests in Grade 1 and 28% were missing their spelling tests in Grade 3. Little's (1988) test of Missing Completely at Random (MCAR) showed that missingness was not completely at random, $\chi^2(36) = 47.07$, $p < .01$. The covariate-dependent missingness (CDM) or missing at random assumption, an extension of Little's MCAR test when covariates are present, was further conducted (Li, 2013). Little's CDM test indicated that adding EL status as a covariate can pass the CDM test, $\chi^2(36) = 47.07$, $p = .10$. Thus, although data of children's spelling tests was not MCAR, the missingness can be viewed as CDM given the children's EL status (Li, 2013), and therefore EL status was added as a covariate to CLPMs. The coefficients for EL status in all CLPMs where it was included as a covariate are reported in Table A5 in the Appendix. To account for missingness while maximizing the available data, pairwise deletion was used for correlation analysis and the Full Information Maximum Likelihood (FIML) estimator was used to handle missing values for CLPMs.

Model fit statistics for CLPMs run for RQs 2 and 3 were not available due to the CLPMs being just-identified with two time points. For multi-group analysis where pathways were constrained to be equal, model fit was available and assessed through several indices, including chi-square statistics, the comparative fit index

(CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residuals (SRMR). For good model fit, RMSEA values should be less than .08, CFI and TLI values should be equal to or higher than .95, and SRMR should be .05 or lower (Hu & Bentler, 1999; Kline, 2015).

Results

Research Question 1: How do spelling skills in English and Spanish of Spanish-English bilingual children change from Grade 1 to Grade 3?

Descriptive statistics for children's correctness and text distance scores in Grades 1 and 3 are presented in Table 1. Children's mean performance measured as correctness scores for English spelling in Grade 1 was quite low (M = 1.98, SD = 3.37) but showed improvement in Grade 3 with a mean of 8.76 (SD = 5.62). Children's mean performance for Spanish spelling using correctness scores was 2.71 (SD = 2.40) and showed slight improvement to 5.34 points, on average (SD = 3.13). Text distance scores also show children's improvement in both English and Spanish spelling from Grade 1 to Grade 3 since higher scores indicate more distance (or more incorrectness) between attempted and target spelling. Children's mean text distance score for English decreased from 0.49 (SD = 0.21) to 0.19 (SD = 0.18) and decreased from 0.41 (SD = 0.23) to 0.24 (SD = 0.16) for Spanish.

Histograms were generated to visually examine and compare the distribution of different spelling scores (see Figure 1 for correctness and Figure 2 for text distance scores). Grade 1 English correctness scores did not meet the assumption of normality, as indicated by its high kurtosis (9.72), whereas the other English correctness scores appeared normal based on histograms, skewness, and kurtosis (Byrne, 2013; Kline, 2015). To account for the non-normality of Grade 1 English correctness scores, the Maximum-Likelihood estimator with robust standard errors (MLR) was used for fitting CLPMs to answer subsequent research questions using correctness scores for spelling. Text distance scores for English and Spanish spelling in both grade levels had skewness and kurtosis within acceptable ranges and the Maximum-Likelihood (ML) estimator was used when fitting CLPMs using text distance scores. Overall, comparing the histograms for each spelling test with correctness and text distance scores, we see that the distribution of children's spelling scores differs based on which scoring method is used.

Research Question 1a: Do Spanish-English bilingual children's spelling skills differ by instructional context?

Descriptive statistics (means and SDs) and *t*-test results for each spelling score by instructional program are presented in Table 2. For correctness scores, significant differences were found between instructional programs except for Grade 3 English spelling, $t(149) = 0.61$, $p = .27$. The mean score of English spelling in Grade 1 for children in dual immersion programs was significantly lower than to those in English immersion programs, $t(187) = 6.00$, $p < .001$. Results for Spanish spelling indicate significant differences between instructional programs in Grade 1($t(187) = -6.07$, $p < .001$) and a more substantial gap in Grade 3 ($t(149) = -9.30$, $p < .001$), with children in dual immersion programs consistently outperforming those in English immersion programs. Similar results were found for text distance scores although the difference between instructional programs for Grade 3 English spelling was more pronounced compared to when correctness scores were used, $t(149) = -1.43$, $p = .08$. Children in English immersion programs did better on Grade 1 English spelling ($t(187) = -5.78$, $p < .001$), whereas children in dual immersion programs did better on Spanish spelling in both grade levels (Grade 1: $t(187) = 8.375$, $p < .001$; Grade 3: $t(149) = 11.15$, $p < .001$).

Next, bivariate correlations between spelling scores are presented in Table 3. For correctness scores, children's performance in Grade 1 English spelling had a moderate relation with their English spelling score in Grade 3 ($r = .48$, $p < .001$) but did not have a significant relation with their Spanish spelling score in both Grade 1 and Grade 3 ($r = .13$, $p = .09$; $r = -.001$, $p = .99$, respectively). Spanish spelling scores in Grade 1 had a strong relation with Spanish spelling in Grade 3 ($r = .70$, $p < .001$) and a weak relation with English spelling in Grade 3 ($r = .36$, $p < .001$). Children's English and Spanish spelling scores in Grade 3 had a significant moderate relation ($r = .39$, $p < .001$).

Bivariate correlations for text distance scores revealed similar patterns to correctness scores, with only minor differences in the strength of the relation. However, Grade 1 English text distance scores had a significant weak relation with Grade 1 Spanish text distance scores ($r = .20$, $p < .01$) compared to the nonsignificant relation found with correctness scores ($r = .13$, $p = .09$), although the difference between the coefficients were small in magnitude. Correctness and text distance scores had a strong negative relation with each other across English and Spanish spelling tests in both grade levels ($|rs| > .75$, $p < .001$) indicating the opposite directionality of the scores.

Research Question 2: How are Spanish and English spelling interrelated over time from Grade 1 to 3?

To examine how Spanish and English spelling are interrelated over time from Grade 1 to 3, a CLPM with correctness spelling scores was fit (see Figure 3). A strong autoregressive effect was found for Spanish spelling scores ($B = .71$, $p < .001$) and a moderate autoregressive effect was found for English spelling scores from Grade 1 to 3 ($B = .44$, $p < .001$). A weak significant positive relation was found from Grade 1 Spanish to Grade 3 English spelling ($B = .28$, $p < .001$) and a weak significant negative relation was found from Grade 1 English to Grade 3 Spanish spelling ($B = -.14$, $p < .001$). The coefficients for EL status for this model and all CLPMs presented below where it was included as a covariate are reported in Table A5 in the Appendix.

Research Question 3: Does the relation between Spanish and English spelling differ based on which spelling scoring method is used?

Next, a CLPM with text distance spelling scores was fit to examine whether the relation between Spanish and English spelling over time differs by which spelling scoring method is used (see Figure 4). Similar to when correctness scores were used, a strong autoregressive effect was found for Spanish spelling scores ($B = .73$, $p < .001$) and a moderate autoregressive effect was found for English spelling scores from Grade 1 to 3 ($B = .44$, $p < .001$). A weak significant positive relation was found from Grade 1 Spanish to Grade 3 English spelling ($B = .25$, $p < .01$). In contrast to correctness scores, no significant relation was found between Grade 1 English to Grade 3 Spanish spelling when text distance scores were used ($B = -.07$, $p = .17$).

Research Question 4: Does the relation between Spanish and English spelling differ based on which instructional program children are in?

Lastly, multi-group analyses were conducted using both correctness and text distance scores to see whether the relation between Spanish and English spelling differs depending on which instructional program children attend. A multi-group analysis by instructional program using correctness scores where all pathways were constrained to be equal was fitted first (see Figure 5). Model fit was excellent for this most parsimonious model, χ^2 ($df = 6$) = 6.98, $p = .32$; CFI = .99; TLI = .98; RMSEA = .04 [.00-.14]; SRMR = .05.A well-fitting equality-constrained model indicates that the relations can be considered invariant across groups (Kline, 2015). To be conservative, we also estimated a model in which one pathway (G1 Spanish → G3 Spanish) was freed to vary, as this path showed the smallest between-group difference in magnitude and significance in a fully freed model. The chi-square difference test using the Satorra–Bentler scaled correction for MLR indicated no significant improvement in model fit ($\Delta\chi^2 = 2.17$, $p = .14$). Therefore, the fully constrained model was retained as the best-fitting and most parsimonious model, and no additional models were estimated. For children in both English and Dual immersion instructional programs, a significant moderate autoregressive effect was found for English and Spanish spelling (English: $B = .49$, $p < .001$; Spanish: $B = .48$, $p < .001$). The relation from Grade 1

Spanish to Grade 3 English was weak but significant ($B = .26$, $p < .01$) and the relation from Grade 1 English to Grade 3 Spanish was nonsignificant ($B = .06$, $p = .38$).

When a multi-group analysis with all pathways constrained to be equal across instructional programs was fit using text distance, model fit was acceptable for some indices but not great, χ^2 ($df = 6$) = 14.75, $p < .05$; CFI = .96; TLI = .86; RMSEA = .12 [.04-.20]; SRMR = .07. To determine whether model fit can significantly improve, multi-group analysis with all pathways free to vary was conducted again in text distance's original scale. Because there were no specific a priori hypotheses about which relations would vary across instructional programs, the decision to free or constrain paths was data-driven, guided by the largest observed differences in magnitude and significance. A series of models was then estimated, sequentially adding constraints to each pathway and evaluating model fit using chi-square difference tests. Model fit statistics and results are summarized in Table 4, where Model 1 only constrained the pathway from Grade 1 English to Grade 3 Spanish to be equal and the last model constrained all pathways to be equal. No significant difference was found between models up to Model 3 where all pathways except for the relation from Grade 1 Spanish to Grade 3 English was constrained, $\Delta\chi^2 = 2.57$, $p = .11$. When the relation from Grade 1 Spanish to Grade 3 English was also constrained in Model 4, chi-square difference test indicated a significant difference from Model 3, $\Delta\chi^2 = 4.15$, $p < .05$. Thus, Model 3 that demonstrated the best relative fit was retained as the final model.

The final multi-group analysis model by instructional programs using text distance scores is presented in Figure 6. As shown above, model fit was good, χ^2 ($df = 5$) = 10.60, $p = .06$; CFI = .97; TLI = .89; RMSEA = .10 [.00-.19]; SRMR = .07. A significant moderate autoregressive effect for English spelling ($B = .43$, $p < .001$) and Spanish spelling ($B = .38$, $p < .001$) was found for children regardless of instructional program. A significant weak relation was found from Grade 1 English to Grade 3 Spanish ($B = .21$, $p < .01$). Only the pathway from Grade 1 Spanish to Grade 3 English was freed to vary across instructional programs. For children in the English immersion program, no significant relation was found from Grade 1 Spanish to Grade 3 English ($B = .16$, $p = .06$) whereas a moderate significant relation was found for children in the Dual immersion program ($B = .47$, $p < .01$).

Discussion

The aim of this study was to enhance our understanding of Spanish-English emergent bilingual children's spelling development by examining the relation between their spelling skills in both languages and how these skills interrelate over time. The study also explored the utility of a nonbinary scoring measure for assessing

Spanish spelling and whether the use of a more nuanced score shows different relations. Lastly, we examined whether the language of instruction (based on the instructional program children attended; English immersion or dual immersion program) influences spelling development and the interrelation between Spanish and English. Understanding these relations is an important endeavor for informing effective instructional strategies and supporting Spanish-English bilingual children's spelling development.

English and Spanish Spelling Development and the Role of Instructional Program

Children in both English immersion and dual immersion programs showed improvement in their spelling in both languages from Grade 1 to 3 with the specific instructional context influencing their progress. Notably, using correctness scores for English spelling in Grade 1 resulted in a low mean score, with nearly half of the children scoring zero on the dichotomously scored test. However, when text distance scores were used for Grade 1 English spelling, this floor effect was mitigated as text distance scores give partial credit for students' developing knowledge. When spelling scores were compared by instructional program, a significant difference was found for English spelling scores in Grade 1 but not in Grade 3. Contrary to our hypothesis, children in English immersion programs did not outperform their peers in dual immersion programs in English spelling after two additional years of instruction. In contrast, the results for Spanish spelling aligned with our hypothesis, showing a widening gap where dual immersion children were significantly outperforming their peers in English immersion programs in Grade 3. In the U.S. context, children, including those in dual immersion programs, are widely exposed to English outside of school, allowing them to continue developing their English spelling skills even with comparatively less formal English instruction. In contrast, opportunities to engage with written Spanish can be often more limited beyond the home environment or neighborhood depending on communities, making Spanish literacy instruction at school particularly important for continued development. This pattern aligns with Ellis's (2002) input-driven theory, which posits that language learning is closely tied to the amount and type of input learners receive across contexts. The observed pattern in the current study where Grade 3 children in dual immersion programs performed comparably to their peers in English immersion programs on English spelling and outperformed on Spanish spelling suggests that environmental exposure and instructional input play roles in bilingual children's spelling development.

Relations Between English and Spanish Spelling Using Correctness Scores

English and Spanish spelling correctness scores were related over time, and importantly, a significant bidirectional relation was found between Spanish and English spelling, over and above autoregressive effects.

Stronger autoregressive effects were found for Spanish spelling indicating that prior spelling has a more prominent role in later development for Spanish than English spelling. Grade 1 Spanish spelling positively predicted Grade 3 English spelling, suggesting a positive crosslinguistic transfer effect of Spanish spelling skills on English spelling development, such that spelling skills in Spanish supports spelling skills in English. Although prior studies suggested a temporary negative transfer for specific English spelling errors where Spanish influence was identified (Fashola et al., 1996; Wolters & Kim, 2023; Zutell & Allen, 1988), our results indicate that, when examined at the broader level of overall spelling proficiency, Spanish spelling facilitates English spelling development even in the presence of cross-linguistic spelling errors. These findings contribute to a growing understanding of cross-linguistic influence in bilingual literacy, showing spelling learned in one language can facilitate their spelling development in another language.

Conversely, Grade 1 English spelling negatively influenced Grade 3 Spanish spelling, potentially due to negative cross-linguistic transfer from the discrepancies between the two languages. Previous studies have also found Spanish spelling errors in children's writing due to English influence (Bahr et al., 2015; Wolters & Kim, 2023). According to Cummins' (1979, 1992) Linguistic Interdependence Theory, positive linguistic transfer requires a certain level of proficiency. The majority of children in our study could not spell more than one English word correctly in Grade 1, and their low proficiency in English spelling may have prevented positive transfer from occurring. Furthermore, because English has a deeper orthography with less consistent phoneme–grapheme mappings (Orthographic Depth Hypothesis; Katz & Frost, 1992; Seymour et al., 2003), transfer from English to Spanish may be more likely to result in interference than facilitation, whereas transfer from Spanish to English—with its more regular and transparent orthographic correspondences—may be more beneficial. Given that irregularities in English spelling make it more challenging to learn than Spanish (Figueredo, 2006; Gaintza & Goikoetxea, 2016; Katz & Frost, 1992), language-general skills acquired from Spanish spelling may be more readily transferable to English spelling. On the other hand, most English spelling rules that children have to learn due to irregular phoneme-grapheme mappings will not be applicable to Spanish spelling.

Relations Between English and Spanish Spelling Using Text Distance Scores

An important aspect of the current study was the examination of the relation between English and Spanish spelling with different evaluative approaches, using text distance scores, which capture children's spelling abilities in a more nuanced way. Overall, the results were similar to those obtained with correctness scores, except that the previously significant negative effect of Grade 1 English spelling on Grade 3 Spanish spelling was not observed when using text distance scores. Focusing on what each score represents can help explain the different results. Both correctness and text distance scores tap into children's spelling ability. However, whereas the former reflects accuracy at the word level, the latter credits partial accuracy and provides a fine-grained representation of children's developing spelling skills. Aligned with these conceptual differences, the two scoring approaches differed in their distributional properties (see Figures 1 and 2), which affects the amount of information each provides. For instance, histograms of Grade 1 English spelling scores displayed a positive skew with a floor effect suggesting that this scoring method is limited in the ability to detect variability among emerging spellers (Caravolas et al., 2001). The more fine-grained text distance scores had a normal distribution, capturing varying degrees of correctness and partial knowledge leading to different outcomes. Overall, the results support the DIEW's dynamic relations as a function of measurement (Kim, 2020; Kim & Graham, 2022) —the cross-linguistic relations between English and Spanish spelling varied depending on how spelling was evaluated, suggesting that observed relations between literacy skills can differ as a function of measurement. The results highlight the importance of carefully considering the choice of measurement and using nuanced measurement approaches to fully understand children's spelling development.

Cross-Linguistic Spelling Relations by Instructional Programs

We also examined whether the interrelation between Spanish and English spelling varied by instructional program. Using correctness scores, the relations between spelling tests did not differ by instructional program. In both English immersion and dual immersion programs, Grade 1 English spelling did not significantly predict Grade 3 Spanish spelling, whereas Grade 1 Spanish spelling significantly predicted Grade 3 English spelling. However, when text distance scores were used, model fit was not as good with all pathways constrained to be equal across instructional groups, suggesting potentially different relations by group. Overall, spelling development did not differ significantly by instructional program, as the final model constrained most pathways to be equal except for one (i.e., Grade 1 Spanish to Grade 3 English). This might be due to the similarities between the two languages and the common developmental pathway for Spanish and English spelling found in prior research (Defior & Serrano, 2005; Rubin & Carlan, 2005).

The relation from Grade 1 Spanish to Grade 3 English text distance scores was the only differing between groups in multigroup analysis by instructional program. Aligned with our hypothesis, Grade 1 Spanish had a significant effect on Grade 3 English only for children in dual immersion programs but not for children in English immersion programs. We can draw on Cummins' (1979, 1992) Linguistic Interdependence Hypothesis again which posits that cross-linguistic transfer can be limited when one's language proficiency does not reach a certain threshold, which is the case for the Spanish spelling of emergent bilingual children in English immersion programs in this study (as indicated by their low mean Spanish spelling score; M = 1.53 out of 15). This is further explained by the input-driven theory (Ellis, 2002), which posits that language learning is closely aligned with the input children receive, including the language of instruction. Children in English immersion programs, who receive most of their instruction in English, may not have had sufficient development in their Spanish spelling by Grade 1 for it to play a positive role in English spelling development. This finding is in line with prior research that found moderating effects of instructional programs on linguistic transfer (Kuo et al., 2017; San Francisco et al., 2006; Wolters & Kim, 2023) and underscores the importance of instructional input in shaping Spanish-English bilingual children's spelling development.

Limitations

This study has several limitations worth noting. First, the study was only able to include two time points to examine change (Grades 1 and 3), which limits the ability to capture the full trajectory of spelling development in emergent bilingual children. With only two time points, we were unable to obtain fit statistics for the CLPMs, which can be used to assess the model's adequacy. In addition, we were unable to fit RI-CLPMs (Hamaker et al., 2015) to separate within- and between-person variance because the model requires at least three waves. We interpreted cross-lagged paths as reflecting associations between earlier and later spelling, after adjusting for prior performance, rather than as evidence of causal effects. Despite this limitation, the findings yielded novel findings (especially in examining different scoring methods) and important insights into the developmental relations between Spanish and English spelling and contribute significantly to the limited body of research on bilingual children's overall spelling development. Future studies that include additional time points for the CLPMs could offer a more comprehensive examination of developmental pathways and yield more robust insights into the relations between spelling skills in both languages.

Second, there was attrition in the data from Grade 1 to Grade 3 due to the longitudinal nature of the study. We accounted for missing data using the FIML estimator and added EL status as a covariate to address the nature of missingness of the data, data with less longitudinal attrition would have been ideal. Next, as one of the main goals of the current study was to examine the appropriateness of using text distance scores itself, we did not include other nonbinary scoring approaches that have been examined in prior studies to take a more nuanced approach to scoring spelling (e.g., Correct Letter Sequence; see Lee and Al Otaiba [2017], for a review of scoring methods). Future research should compare the automated text distance scores with human-coded, nonbinary scoring methods to further evaluate their interpretability and educational utility. Also, the mean scores on the spelling tests were relatively low, particularly for the English spelling test in Grade 1, leading to a floor effect when using correctness scores. Administering easier spelling tests might capture a wider range of spelling abilities and provide more nuanced insights into the development of spelling skills.

Lastly, although EL status was included as a covariate in the analyses to address patterns of missingness, some differences in EL designation by instructional program remains and could represent a potential confound, even when statistically controlled. This limitation reflects the close alignment between program type and language background in bilingual education contexts. Additionally, information on children's home language environments was not available, which could have provided further insight into the influence of home literacy environment. It is important to note that our primary focus in the current study was on the instructional program, or language of instruction, rather than EL designation or home language, and all participating children were bilingual learners who varied in their degree of English and Spanish proficiency. Future research with larger and more balanced samples, along with information on home language environments, would help disentangle the effects of EL status, instructional program, and home language environments to better understand how these factors influence bilingual children's spelling development.

Implications

Despite the aforementioned limitations, there are insightful implications that can be drawn from this study. First, this study highlights the importance of understanding the longitudinal relation between Spanish and English spelling in emergent bilingual children and provides insights to inform effective instructional strategies and support spelling development. Our findings suggest that Spanish spelling proficiency positively influences English spelling development, while early English spelling may impede Spanish spelling due to cross-linguistic differences. Early instruction in Spanish in dual immersion programs can support later English spelling for bilingual children, emphasizing the importance of strengthening L1 literacy skills to scaffold L2 development. To prevent negative transfer, bilingual children can benefit from receiving explicit instruction on the differences between Spanish and English spelling. Furthermore, children receiving more Spanish instruction outperformed peers in Spanish spelling suggesting that spelling instruction is necessary for balanced biliteracy development. Future studies with children at more advanced phases of development could determine if these findings hold true across different developmental phases.

Next, our study underscores the importance of using nuanced assessment methods, such as text distance scores for spelling, to gain a more detailed understanding of bilingual children's spelling development. To our knowledge, the current study is the first to demonstrate that text distance scores can be used for Spanish spelling. We find one differing interrelation between Spanish and English spelling when this nonbinary scoring method is used. Instructionally, text distance offers a more asset-based approach by recognizing partial knowledge and progress that may be overlooked by binary scoring, which is especially important for bilingual children who find it difficult to produce fully correct English spellings yet. Binary scoring may obscure growth, provide limited feedback for teachers, and discourage learners. Future research should continue to explore the use of nonbinary scoring measures for bilingual children to provide a more comprehensive understanding of children's spelling abilities and nuanced ways in which their literacy skills in one language support or hinder their development in the other.

Finally, we also found some developmental differences depending on which instructional program children attend, emphasizing the importance of considering the language of instruction when examining emergent bilingual's spelling development. We found that children perform better on spelling tests in the language they receive more instruction in, and those in dual immersion programs may experience more cross-linguistic transfer. Bilingual children would benefit from receiving input or instruction in both languages for transfer between two languages to occur. This is especially important for the home language, as opportunities to develop spelling skills outside the classroom may be limited. Given the inherent challenges of learning to spell in two languages, it is crucial to explore how spelling skills in each language develop and interrelate over time.

References

Authors (2023)

Authors (2024)

Author (2025)

Apel, K., Wilson-Fowler, E. B., Brimo, D., & Perrin, N. A. (2012). Metalinguistic contributions to reading and spelling in second and third grade students. *Reading and Writing*, *25*(6), 1283–1305. https://doi.org/10.1007/s11145-011-9317-8

Bahr, R. H., Silliman, E. R., Danzak, R. L., & Wilkinson, L. C. (2015). Bilingual spelling patterns in middle school: It is more than transfer. *International Journal of Bilingual Education and Bilingualism*, *18*(1), 73–91. https://doi.org/10.1080/13670050.2013.878304

Branum-Martin, L., Mehta, P. D., Fletcher, J. M., Carlson, C. D., Ortiz, A., Carlo, M., & Francis, D. J. (2006). Bilingual phonological awareness: Multilevel construct validation among Spanish-speaking kindergarteners in transitional bilingual education classrooms. *Journal of Educational Psychology*, *98*(1), 170–181. https://doi.org/10.1037/0022-0663.98.1.170

Byrne, B. M. (2013). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Routledge. https://www.taylorfrancis.com/books/mono/10.4324/9780203807644/structural-equation-modeling-mplus-barbara-byrne

Caravolas, M., Hulme, C., & Snowling, M. J. (2001). The foundations of spelling ability: Evidence from a 3-year longitudinal study. *Journal of Memory and Language*, *45*(4), 751–774. https://doi.org/10.1006/jmla.2000.2785

Caravolas, M., Kessler, B., Hulme, C., & Snowling, M. (2005). Effects of orthographic consistency, frequency, and letter knowledge on children's vowel spelling development. *Journal of Experimental Child Psychology*, *92*(4), 307–321. https://doi.org/10.1016/j.jecp.2005.08.001

Clemens, N. H., Oslund, E. L., Simmons, L. E., & Simmons, D. (2014). Assessing spelling in kindergarten: Further comparison of scoring metrics and their relation to reading skills. *Journal of School Psychology*, *52*(1), 49–61. https://doi.org/10.1016/j.jsp.2013.12.005

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, *49*(2), 222–251. https://doi.org/10.3102/00346543049002222

Cummins, J. (1992). Bilingualism and second language learning. *Annual Review of Applied Linguistics*, *13*, 51–70. https://doi.org/10.1017/S0267190500002397

Deacon, S. H., Chen, X., Luo, Y., & Ramirez, G. (2013). Beyond language borders: Orthographic processing and word reading in Spanish–English bilinguals. *Journal of Research in Reading*, *36*(1), 58–74. https://doi.org/10.1111/j.1467-9817.2011.01490.x

Deacon, S. H., Kirby, J. R., & Casselman-Bell, M. (2009). How robust is the contribution of morphological awareness to general spelling outcomes? *Reading Psychology*, *30*(4), 301–318. https://doi.org/10.1080/02702710802412057

Defior, S., & Serrano, F. (2005). The initial development of spelling in Spanish: From global to analytical. *Reading and Writing*, *18*, 81–98. https://doi.org/10.1007/s11145-004-5893-1

Dixon, L. Q., Zhao, J., & Joshi, R. M. (2010). Influence of L1 orthography on spelling English words by bilingual children: A natural experiment comparing syllabic, phonological, and morphosyllabic first languages. *Learning Disability Quarterly*, *33*(3), 211–221. https://doi.org/10.1177/073194871003300309

Ehri, L. C. (1987). Learning to read and spell words. *Journal of Reading Behavior*, *19*(1), 5–31. https://doi.org/10.1080/10862968709547585

Ehri, L. C. (2000). Learning to read and learning to spell: Two sides of a coin. *Topics in Language Disorders*, *20*(3), 19–36.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*(2), 143–188. https://doi.org/10.1017/S0272263102002024

Fashola, O. S., Drum, P. A., Mayer, R. E., & Kang, S.-J. (1996). A cognitive theory of orthographic transitioning: Predictable errors in how Spanish-speaking children spell English words. *American Educational Research Journal*, *33*(4), 825–843. https://doi.org/10.3102/00028312033004825

Figueredo, L. (2006). Using the known to chart the unknown: A review of first-language influence on the development of English-as-a-second-language spelling skill. *Reading and Writing*, *19*(8), 873–905. https://doi.org/10.1007/s11145-006-9014-1

Gaintza, Z., & Goikoetxea, E. (2016). Spelling instruction in Spanish: A comparison of self-correction, visual imagery and copying. *Journal of Research in Reading*, *39*(4), 428–447. https://doi.org/10.1111/1467-9817.12055

Georgiou, G. K., Torppa, M., Manolitsis, G., Lyytinen, H., & Parrila, R. (2012). Longitudinal predictors of reading and spelling across languages varying in orthographic consistency. *Reading and Writing*, *25*(2), 321–346. https://doi.org/10.1007/s11145-010-9271-x

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102–116. https://doi.org/10.1037/a0038889

Helman, L., Bear, D. R., Johnston, F., Invernizzi, M., & Templeton, S. (2013). *Palabras a su paso: El estudio de palabras en acción [Words their Way: The study of words in action]*. Pearson.

Hevia-Tuero, C., Russak, S., & Suárez-Coalla, P. (2023). Spelling errors by Spanish children when writing in English as a foreign language. *Reading and Writing*, *36*(7), 1797–1820. https://doi.org/10.1007/s11145-022-10356-5

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In L. Katz & R. Frost (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 67–84). Elsevier Science Publishers.

Kim, Y.-S., Al Otaiba, S., Folsom, J. S., Greulich, L., & Puranik, C. (2014). Evaluating the dimensionality of first-grade written composition. *Journal of Speech, Language, and Hearing Research*, *57*(1), 199–211. https://doi.org/10.1044/1092-4388(2013/12-0152)

Kim, Y.-S., Apel, K., & Al Otaiba, S. (2013). The relation of linguistic awareness and vocabulary to word reading and spelling for first-grade students participating in response to intervention. *Language, Speech, and Hearing Services in Schools*, *44*, 1–11. https://doi.org/10.1044/0161-1461(2013/12-0013)

Kim, Y.-S. G. (2020). Structural relations of language and cognitive skills, and topic knowledge to written composition: A test of the direct and indirect effects model of writing. *British Journal of Educational Psychology*, *90*(4), 910–932. https://doi.org/10.1111/bjep.12330

Kim, Y.-S. G., & Graham, S. (2022). Expanding the Direct and Indirect Effects Model of Writing (DIEW): Reading–writing relations, and dynamic relations as a function of measurement/dimensions of written composition. *Journal of Educational Psychology*, *114*(2), 215–238. https://doi.org/10.1037/edu0000564

Kim, Y.-S. G., & Schatschneider, C. (2017). Expanding the developmental models of writing: A direct and indirect effects model of developmental writing (DIEW). *Journal of Educational Psychology*, *109*, 35–50. https://doi.org/10.1037/edu0000129

Kim, Y.-S. G., Wolters, A., Mercado, J., & Quinn, J. (2022). Crosslinguistic transfer of higher order cognitive skills and their roles in writing for English-Spanish dual language learners. *Journal of Educational Psychology*, *114*(1), 1–15. https://doi.org/10.1037/edu0000516

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford publications.

Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language Learning*, *57*(1), 1–44. https://doi.org/10.1111/j.1467-9922.2007.00411.x

Kruskal, J. B. (1983). An overview of sequence comparison. In D. Sankoff & J. B. Kruskal (Eds.), *Time warps, string edits, and macromolecules: The theory and practice of sequence comparison* (Vol. 25, pp. 1–44). Addison-Wesley.

Kuo, L.-J., Ramirez, G., de Marin, S., Kim, T.-J., & Unal-Gezer, M. (2017). Bilingualism and morphological awareness: A study with children from general education and Spanish-English dual language programs. *Educational Psychology*, *37*(2), 94–111. https://doi.org/10.1080/01443410.2015.1049586

Lee, J. A. C., & Al Otaiba, S. (2017). End-of-kindergarten spelling outcomes: How can spelling error analysis data inform beginning reading instruction? *Reading & Writing Quarterly*, *33*(3), 226–238. https://doi.org/10.1080/10573569.2016.1165639

Levesque, K. C., Breadmore, H. L., & Deacon, S. H. (2021). How morphology impacts reading and spelling: Advancing the role of morphology in models of literacy development. *Journal of Research in Reading*, *44*(1), 10–26. https://doi.org/10.1111/1467-9817.12313

Li, C. (2013). Little's Test of Missing Completely at Random. *The Stata Journal*, *13*(4), 795–809. https://doi.org/10.1177/1536867X1301300407

Lindner, A. L., Wijekumar, K., & Joshi, R. M. (2022). English spelling performance in writing samples among Spanish-speaking ELLs. *Journal of Learning Disabilities*, *55*(2), 114–122. https://doi.org/10.1177/0022219420982995

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202.

Lucas, R. E. (2023). Why the Cross-Lagged Panel Model Is Almost Never the Right Choice. *Advances in Methods and Practices in Psychological Science*, *6*(1), 25152459231158378. https://doi.org/10.1177/25152459231158378

MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (Vol. 4967, pp. 50–70). Oxford University Press.

Moats, L. (2009). Knowledge foundations for teaching reading and spelling. *Reading and Writing*, *22*(4), 379–399. https://doi.org/10.1007/s11145-009-9162-1

Moon, Y., & Kim, Y.-S. G. (2023). *Spelling Dystance*. https://spellingdystance.education.uci.edu/

Moreau, E., Yvon, F., & Cappé, O. (2008). Robust similarity measures for named entities matching. *22nd International Conference on Computational Linguistics*, 593–600. https://aclanthology.org/C08-1075.pdf

orsinium. (n.d.). *textdistance: Compute distance between the two texts.* (Version 4.2.2) [Python]. Retrieved May 18, 2022, from https://github.com/orsinium/textdistance

Pérez Cañado, M. L. (2005). English and Spanish spelling: Are they really different? *The Reading Teacher*, *58*(6), 522–530. https://doi.org/10.1598/RT.58.6.3

Rubin, R., & Carlan, V. G. (2005). Using writing to understand bilingual children's literacy development. *The Reading Teacher*, *58*(8), 728–739. https://doi.org/10.1598/RT.58.8.3

San Francisco, A. R., Mo, E., Carlo, M., August, D., & Snow, C. (2006). The influences of language of literacy instruction and vocabulary on the spelling of Spanish–English bilinguals. *Reading and Writing*, *19*(6), 627–642. https://doi.org/10.1007/s11145-006-9012-3

Seymour, P. H., Aro, M., Erskine, J. M., & Network, C. with C. A. A. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174. https://doi.org/10.1348/000712603321661859

Treiman, R. (1993). *Beginning to spell: A study of first-grade children*. Oxford University Press.

Treiman, R. (2017). Learning to spell words: Findings, theories, and issues. *Scientific Studies of Reading*, *21*(4), 265–276. http://dx.doi.org/10.1080/10888438.2017.1296449

Treiman, R., Kessler, B., & Caravolas, M. (2019). What methods of scoring young children's spelling best predict later spelling performance? *Journal of Research in Reading*, *42*(1), 80–96. https://doi.org/10.1111/1467-9817.12241

Treiman, R., Kessler, B., Pollo, T. C., Byrne, B., & Olson, R. K. (2016). Measures of kindergarten spelling and their relations to later spelling performance. *Scientific Studies of Reading*, *20*(5), 349–362. https://doi.org/10.1080/10888438.2016.1186168

Vaughn, S., Cirino, P. T., Linan-Thompson, S., Mathes, P. G., Carlson, C. D., Hagan, E. C., Pollard-Durodola, S. D., Fletcher, J. M., & Francis, D. J. (2006). Effectiveness of a Spanish Intervention and an English Intervention for English-Language Learners at Risk for Reading Problems. *American Educational Research Journal*, *43*(3), 449–487. https://doi.org/10.3102/00028312043003449

Verhoeven, L. (2000). Components in early second language reading and spelling. *Scientific Studies of Reading*, *4*(4), 313–330. https://doi.org/10.1207/S1532799XSSR0404_4

Wawire, B. A., & Kim, Y.-S. G. (2018). Cross-Language Transfer of Phonological Awareness and Letter Knowledge: Causal Evidence and Nature of Transfer. *Scientific Studies of Reading*, *22*(6), 443–461. https://doi.org/10.1080/10888438.2018.1474882

Wolters, A. P., & Kim, Y. G. (2023). Crosslinguistic influence on spelling in written compositions: Evidence from English-Spanish dual language learners in primary grades. *Reading and Writing*. https://doi.org/10.1007/s11145-023-10416-4

Zaretsky, E. (2020). English spelling acquisition by English Language Learners from Spanish-speaking background: The role of cognitive and linguistic resources and L1 reading status. *Cognitive Development*, *55*, 1–12. https://doi.org/10.1016/j.cogdev.2020.100918

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3

Zutell, J., & Allen, V. (1988). The English spelling strategies of Spanish-speaking bilingual children. *TESOL Quarterly*, *22*(2), 333–340. https://doi.org/10.2307/3586941

Tables

Table 1.

Descriptive Statistics for Full Sample

Variables	N	Mean	SD	Min	Max	Skewness	Kurtosis
Correctness Scores							
G1 English	189	1.98	3.37	0	18	2.54	9.72
G3 English	151	8.76	5.62	0	18	0.003	1.61
G1 Spanish	189	2.71	2.40	0	9	0.83	2.67
G3 Spanish	151	5.34	3.13	0	12	0.02	1.86
Text Distance Scores							
G1 English	189	0.49	0.21	0	1	−0.003	2.43
G3 English	151	0.19	0.18	0	1	1.43	5.79
G1 Spanish	189	0.41	0.23	0.09	1	0.82	2.88
G3 Spanish	151	0.24	0.16	0.03	0.73	0.96	3.18

Note. English spelling test included 18 items. Spanish spelling test included 15 items. Text distance scores are composite scores of normalized Damerau-Levenshtein and Cosine Distance scores averaged across all items.

Table 2.

Descriptive Statistics and t-Test Results by Instructional Programs

	English Immersion		Dual Immersion		*t*-test	
Variables	Mean	SD	Mean	SD	*t* (*df*)	*p*-value
Correctness Scores						
G1 English	3.59	4.02	0.85	2.23	6.00 (187)	.000***
G3 English	9.16	5.78	8.56	5.56	0.61 (149)	.27
G1 Spanish	1.53	1.42	3.50	2.59	−6.07 (187)	.000***
G3 Spanish	2.61	1.69	6.66	2.81	−9.30 (149)	.000***
Text Distance Scores						
G1 English	0.39	0.22	0.55	0.18	−5.78 (187)	.000***
G3 English	0.16	0.14	0.20	0.19	−1.43 (149)	.08
G1 Spanish	0.56	0.23	0.30	0.17	8.75 (187)	.000***
G3 Spanish	0.40	0.13	0.17	0.11	11.15 (149)	.000***

Note. English spelling test included 18 items. Spanish spelling test included 15 items. Text distance scores are a composite of normalized Damerau-Levenshtein and Cosine Distance scores. The Bonferroni-adjusted significance threshold for eight comparisons is $p < .00625$.
*** $p < .001$.

Table 3.

Bivariate Correlations between English and Spanish Spelling in Grades 1 and 3

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Correctness							
(1) G1 English	-						
(2) G3 English	0.48***	-					
(3) G1 Spanish	0.13~	0.36***	-				
(4) G3 Spanish	−0.001	0.39***	0.70***	-			
Text Distance							
(5) G1 English	**−0.75*** **	−0.51***	−0.28***	−0.15~	-		
(6) G3 English	−0.39***	**−0.88*** **	−0.31***	−0.37***	0.50***	-	
(7) G1 Spanish	−0.002	−0.35***	**−0.78*** **	−0.69***	0.20**	0.32***	-
(8) G3 Spanish	0.04	−0.33***	−0.59***	**−0.88*** **	0.07	0.29***	0.69***

Note. Bold text denotes correlations between correctness and text distance scores for same spelling tests.
p ~ < .10, *p* *** < .001.

Table 4.

Model Fit Statistics and Chi-square Difference Test between Multi-group CLPMs with Text Distance Scores: Constraints on Pathways Added Cumulatively

Added Constraint	Model Fit Statistics	$\Delta\chi^2$ Difference Test
(1) G3 Spanish on G1 English	χ^2 (df = 3) = 4.92, p = .18; CFI = .99; TLI = .94; RMSEA = .08 [.00-.20]; SRMR = .08	-
(2) G1 Spanish on G3 Spanish	χ^2 (df = 4) = 8.03, p = .09; CFI = .98; TLI = .91; RMSEA = .10 [.00-.20]; SRMR = .07	$\Delta\chi^2$ = 3.11, p = .08
(3) G1 English on G3 English	χ^2 (df = 5) = 10.60, p = .06; CFI = .97; TLI = .89; RMSEA = .10 [.00-.19]; SRMR = .07	$\Delta\chi^2$ = 2.57, p = .11
(4) G1 Spanish on G3 English	χ^2 (df = 6) = 14.75, p < .05; CFI = .96; TLI = .86; RMSEA = .12 [.04-.20]; SRMR = .07	$\Delta\chi^2$ = 4.15, p < .05

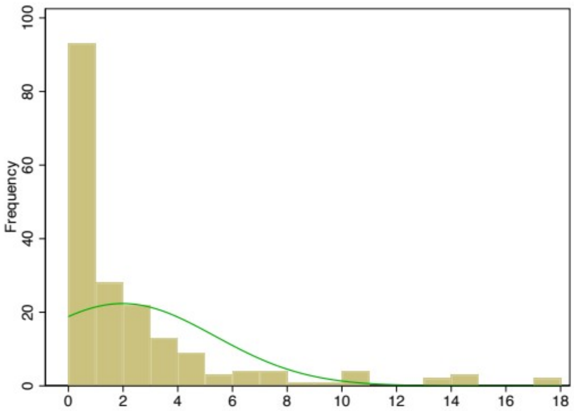
Note. Model fit statistics for each model with added constraints for text distance scores. Constraints on pathways were added in a cumulative manner such that Model (4) constrained all pathways to be equal across instructional programs. EL status was included in all models as a covariate. Final model can be found in Figure 6.

Figures

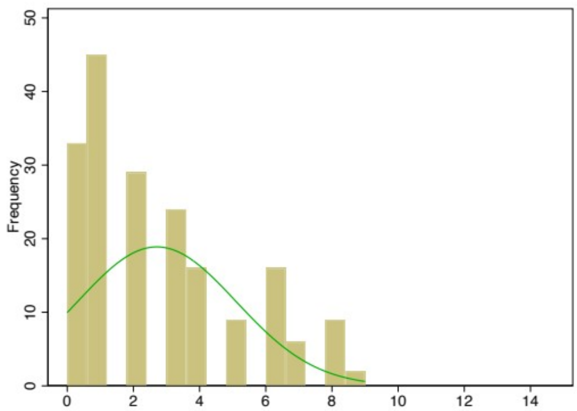
Figure 1.

Histograms of English and Spanish Spelling Correctness Scores in Grade 1 and 3

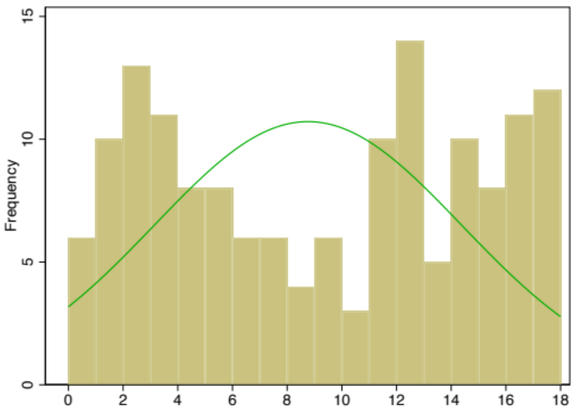
a. Grade 1 English Spelling



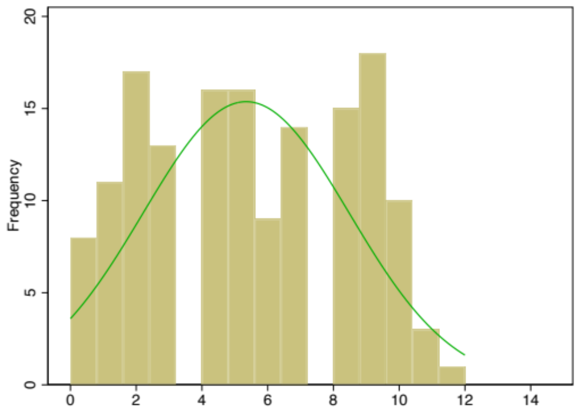
b. Grade 1 Spanish Spelling



c. Grade 3 English Spelling



d. Grade 3 Spanish Spelling

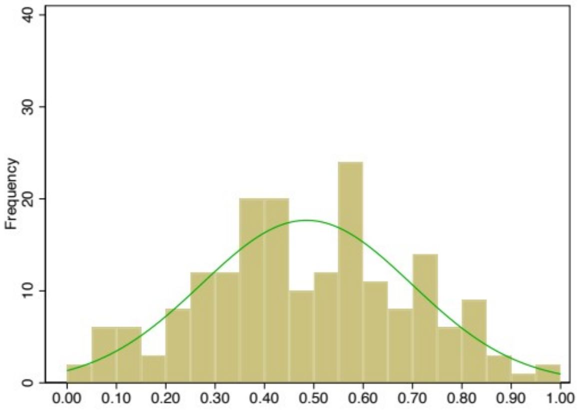


Note. Histograms using correctness scores. Maximum possible score for English spelling test is 18 and 15 for Spanish spelling test.

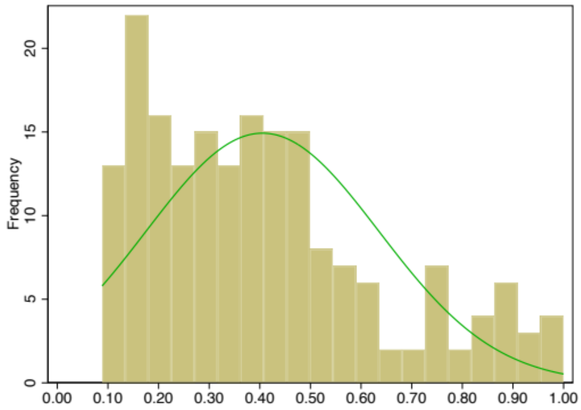
Figure 2.

Histograms of English and Spanish Spelling Text Distance Scores in Grade 1 and 3

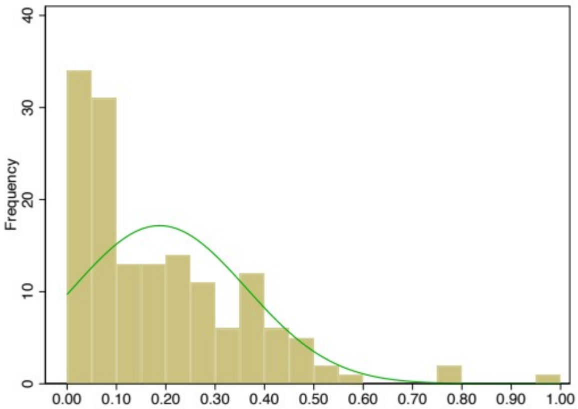
a. Grade 1 English Spelling



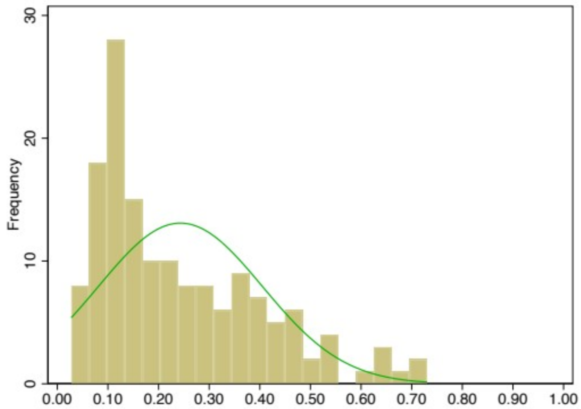
b. Grade 1 Spanish Spelling



c. Grade 3 English Spelling



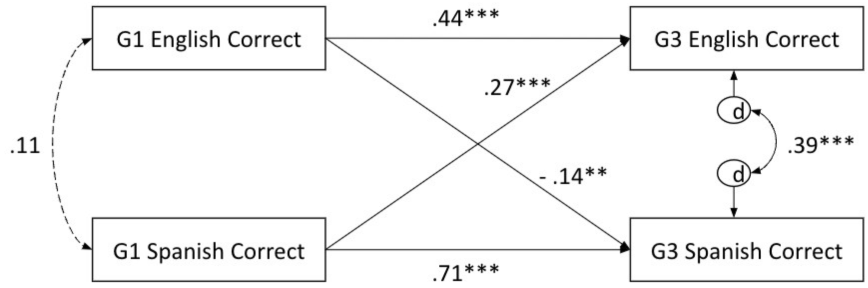
d. Grade 3 Spanish Spelling



Note. Histograms using text distance scores. Range of possible text distance scores for both English and Spanish spelling test is from 0 to 1.

Figure 3.

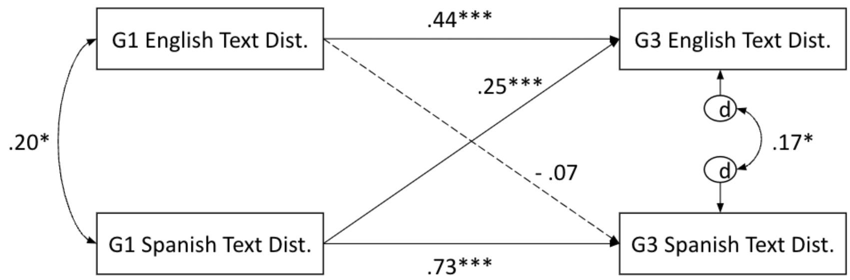
Relations between English and Spanish Spelling Correctness Scores in Grades 1 and 3 for the Full Sample



Note. Standardized coefficients. Dashed lines indicate nonsignificant relations. Disturbance (d) is unexplained variance in the endogenous variables. EL status is included in the model as a covariate but not shown.
p ** < .01, *p* *** < .001.

Figure 4.

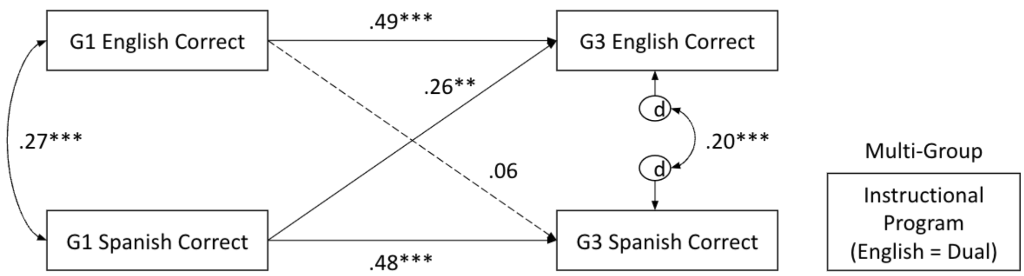
Relations between English and Spanish Text Distance Scores in Grades 1 and 3 for the Full Sample



Note. Standardized coefficients. Dashed lines indicate nonsignificant relations. Disturbance (d) is unexplained variance in the endogenous variables. EL status is included in the model as a covariate but not shown.
p ~ < .10, *p* ** < .01, *p* *** < .001.

Figure 5.

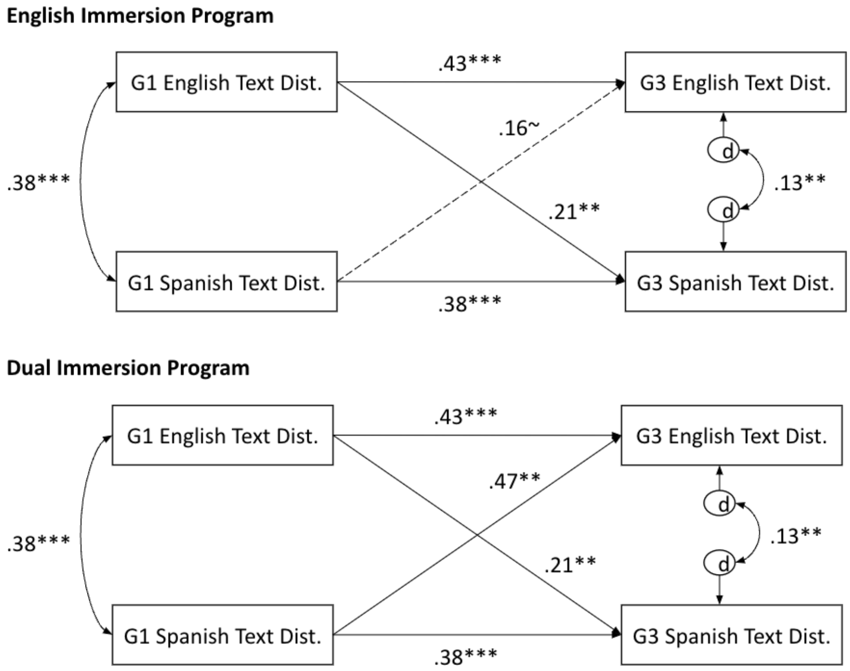
Multi-group Analysis by Instructional Programs with all Pathways Constrained to be Equal: Relations between English and Spanish Correctness Scores in Grades 1 and 3



Note. Standardized coefficients. Dashed lines indicate nonsignificant relations. Disturbance (d) is unexplained variance in the endogenous variables. EL status is included in the model as a covariate but not shown.
p ** < .01, *p* *** < .001.

Figure 6.

Final Multi-group Analysis Model by Instructional Programs: Relations between English and Spanish Text Distance Scores in Grades 1 and 3



Note. All pathways except for G3 English on G1 Spanish are constrained to be equal. Standardized coefficients. Dashed line indicates nonsignificant relations. Disturbance (d) is unexplained variance in the endogenous variables. EL status is included in the model as a covariate but not shown.

p ~ < .10, *p* ** < .01, *p* *** < .001.

Appendix

Table A1.

Number of English Learner (EL) Students by Instructional Program in Grades 1 and 3

Grade – EL Status	English Immersion	Dual Immersion
Grade 1		
Non-English Learner	37 (46 %)	41 (32 %)
English Learner	44 (54 %)	88 (68 %)
Grade 3		
Non-English Learner	26 (46 %)	29 (32 %)
English Learner	30 (54 %)	63 (68 %)

Note. Percentages represent the proportion of students within each instructional program. EL = English Learner.

Table A2.

Number of Students for each Spelling Test by Instructional Program

Grade - Language	English Immersion	Dual Immersion
G1 English	78	111
G3 English	50	101
G1 Spanish	76	113
G3 Spanish	49	102

Table A3.

Word List for English and Spanish Spelling Assessment

#	English	Spanish	Translation
1	hop	el	the
2	ship	pan	bread
3	cute	campos	countryside
4	bump	fresa	strawberry
5	pool	plancha	griddle
6	hang	gigante	giant
7	drive	voy	go
8	girl	aprieto	fix
9	leaf	lápices	pencils
10	phone	autobús	bus
11	bright	hierro	iron
12	making	halla	find
13	catch	extraño	strange
14	hurt	quisiera	I'd like…
15	know	guisante	pea
16	bridge		
17	shopping		

18	sign

Table A4.

Examples of Children's Spellings and their Respective Text Distance Scores

Target Word	Child's Spelling	Normalized Damerau–Levenshtein	Cosine Distance	Composite Score
<girl>	<girl>	0	0	0
<girl>	<gril>	0	0.25	0.125
<girl>	<gl>	0.293	0.5	0.397
<extraño>	<extraño>	0	0	0
<extraño>	<extrano>	0.143	0.143	0.143
<extraño>	<exteno>	0.383	0.429	0.406

Table A5.

Standardized Coefficients on EL Status from CLPM Results

Model	G1 English		G1 Spanish		G3 English		G3 Spanish	
	β	p	β	p	β	p	β	p
Figure 3	−0.26	.000***	−0.04	.60	−0.12	.08~	0.08	.17
Figure 4	0.21	.000***	−0.01	.85	0.10	.18	−0.07	.22
Figure 5								
English Immersion	−0.66	.02*	−0.06	.67	−0.28	.18	−0.09	.53
Dual Immersion	−0.29	.07~	−0.30	.17	−0.22	.27	0.27	.07~
Figure 6								
English Immersion	0.46	.03*	−0.13	.55	0.32	.04*	0.25	.21
Dual Immersion	0.28	.09~	0.30	.05*	0.07	.74	−0.28	.02*

Note. Standardized coefficients and p-values of children's EL status at G1 for each figure.

p ~ < .10, p * < .05, p *** < .001.