Dynamic Assessment for Identifying Spanish-Speaking English Learners' Risk for Mathematics Disabilities: Does Language of Administration Matter?

Submitted to the *Journal of Learning Disabilities* on October 9, 2019

Eunsoo Cho[1]

Lynn S. Fuchs[2]

Pamela M. Seethaler[2]

Douglas Fuchs[2]

Donald L. Compton[3]

Michigan State University[1], Vanderbilt University[2], Florida State University[3]

This research was supported by Award Number R324A090039 from the U.S. Department of Education as well as by Award Number 2 P20 HD075443, 2 R01 HD053714, and Core Grant Number HD15052 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development to Vanderbilt University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the U.S. Department of Education or the Eunice Kennedy Shriver National Institute of Child Health & Human Development or the National Institutes of Health.

Abstract

We examined dynamic assessment's (DA's) added value over traditional assessments for identifying Spanish-speaking English learners' (ELs') risk for developing mathematics disabilities, as a function of the language of test administration (English vs. Spanish), type of math outcome, and EL's language dominance. At the start of 1st grade, ELs (*N* = 368) were randomly assigned to English-DA or Spanish-DA conditions, were assessed on static mathematics measures and domain-general (language, reasoning) measures in English, and completed DA in their assigned language condition. At year's end, they were assessed on calculation and word-problem solving outcomes in English. Results from multi-group path models indicated that Spanish-DA mitigates the impact of ELs' language dominance on DA performance. Moreover, ELs' language dominance moderated DA's predictive validity differentially depending on DA language and type of outcome. Spanish-DA showed higher predictive validity in Spanish-dominant ELs than English-dominant ELs when predicting calculations but not word-problem solving. English-DA was predictive for both outcomes, regardless of ELs' language dominance.

Keywords: dynamic assessment, English learners, language, mathematics disabilities, screening

Dynamic Assessment for Identifying Spanish-Speaking English Learners' Risk for Mathematics Disabilities: Does Language of Administration Matter?

While the number of English learners (ELs) has been growing at a rapid rate in the U.S., comprising approximately 10% of the school population (McFarland et al., 2018), they are also one of the lowest achieving subgroups in mathematics, with mean performance only slightly above the basic level on the NAEP (National Center for Education Statistics, 2017). More troubling is the pervasive achievement gap between ELs and non-ELs. The prevalence of mathematics difficulties is much higher in ELs than in non-ELs: The percentage of ELs (47%) performing below the basic level is more than twice that of non-ELs (17%). Because mathematics difficulties exacerbate over time (Morgan, Farkas, Wu, 2009; National Mathematics Advisory Panel, 2008), there is an urgent need to develop early screening methods to identify ELs who are at risk for developing mathematics disabilities so that timely supplemental intervention may proceed.

Despite advances in prevention and early identification of learning disabilities (LD) with the introduction of response to intervention (RTI; Fuchs & Fuchs, 2006), LD identification for ELs is still 3-4 years delayed. Prior research indicates ELs (racial and linguistic minorities) are less likely to be identified as having disabilities and receive special education services than their counterparts during early childhood (Morgan, Farkas, Hillemeier, & Maczuga, 2012) and throughout elementary school years (Morgan et al., 2015; Zehler, Fleischman, Hopstock, Pendzick, & Stephenson, 2003). A more complex picture is provided by other studies reporting underrepresentation of ELs in special education at young ages but overrepresentation at older grades (Hibel & Jasper, 2012; Samson & Lesaux, 2009). This phenomenon may be due to teachers' propensity to not identify ELs as at risk compared to native English speakers because they presume difficulties are due to limited English proficiency (Limbos & Geva, 2001).

Early identification of ELs with mathematics disabilities is challenging because young ELs' low performance on a mathematics test may occur due to several reasons (e.g., Wagner, Francis, & Morris, 2005). First, ELs may have mathematics competencies but perform poorly on tests due to the language demands of the items on those tests, particularly with word problems (Abedi & Lord, 2001; Martiniello, 2009). Second, ELs may initially struggle with mathematics due to limited proficiency in English, which is their second language, or a lack of environmental supports that facilitate numeracy development, which is often prevalent in low socio-economic backgrounds (e.g., Anders et al., 2012; Magnuson, Meyers, Ruhm, & Waldfogel, 2004). Third, low test scores may accurately depict ELs' true mathematics competence. The first two scenarios, which often co-occur, represent false positive cases in which ELs are inaccurately identified as LD based on low mathematics performance when, in fact, environmental factors or limited English proficiency are masking ELs' mathematics learning abilities. These students may soon catch up as they receive sound mathematics instruction in school and as they acquire English proficiency.

Disentangling various sources of low mathematics performance to distinguish ELs with true risk for mathematics disabilities (true positives) from ELs whose mathematics difficulties result from extraneous factors (false positives) is critical because inaccurate identification of at-risk status and mismatched placements have undesirable consequences at the school and student levels. High false positive rates increase the cost of multi-tiered support systems due to unnecessary expenditures for supplementary intervention (Tier 2) to students who otherwise do well with Tier 1 classroom instruction. At a student level, false positives may result in negative consequences associated with receiving an at-risk label. Moreover, misplacement of students in

Tier 2 can impede their learning (e.g., Balu, Zhu, Doolittle, Schiller, Jenkins, & Gersten, 2015), because instruction may be inappropriately slow-paced with less content coverage (e.g., Oakes, 2005).

Dynamic Assessment

One method to address this longstanding challenge of early identification of ELs at risk for LD is dynamic assessment (DA). DA is a collection of assessment procedures in which structured instruction is provided as part of the testing process to measure how well a student can learn with help (Caffrey, Fuchs, & Fuchs, 2008; Grigorenko & Sternberg, 1998; Grigorenko, 2009; Wagner & Compton, 2012), a developmental phase referred to as the *zone of proximal development* (Vygotsky, 1962). Traditional static assessments measure students' independent performance tapping only the product of learning. This inflates false positive rates because static assessments often show floor effects in young children, many of whom do not perform well because they have insufficient learning opportunities (e.g., Catts, Petscher, Schatschenider, Bridges, & Mendoza, 2009). DA addresses this problem by directly *teaching* the skills assessed on the test items and measuring students' learning in response to that instruction. This permits differentiation between poor performances of students due to lack of prior learning opportunities (false positive) from learning deficits (true positive). In fact, DA has been shown to improve the precision of early identification of risk for LD (Cho, Compton, Fuchs, Fuchs, & Bouton, 2014; Cho, Compton, & Josol, 2019; Compton et al., 2010; Seethaler, Fuchs, Fuchs & Compton, 2012; Gellert & Elbro, 2018; O'Connor & Jenkins, 1999; Petersen, Gragg, & Spencer, 2018).

DA's validity for predicting mathematics outcomes has been examined in several studies with elementary students. DA involving working memory training predicts concurrent and later math outcomes at third through fifth grade (Stevenson, Bergwerff, Heiser, & Resing, 2014;

Swanson & Howard, 2005). Other studies extend that work by focusing the DA's content on mathematics while considering DA's contribution beyond static mathematics screeners and domain-general cognitive assessments. Fuchs, Fuchs, Compton, et al. (2008), whose DA teaches novel mathematics content to third graders, demonstrated DA's incremental validity when predicting response to intervention. As in Seethaler et al. (2012), whose DA relies on math content at first grade, DA was one of the three strongest predictors of the year-end calculation outcome, followed by brief (Quantity Discrimination; Chard et al., 2005) and extended (Test of Early Mathematics Achievement; Ginsburg & Baroody, 2003) math screeners. DA was also the strongest predictor of word-problem (WP) outcomes. While these studies show promise for DA in screening for later mathematics disabilities, little is known about how DA works for ELs.

Whereas a few recent studies have examined DA's validity for predicting reading development in ELs (e.g., Petersen & Gillam, 2015; Peterson, Gragg, & Spencer, 2018), we identified only one prior study that examined DA's utility in predicting later mathematics outcomes in ELs. Authors (2016) examined whether the predictive value of DA using equation-solving tasks at the start of first grade differs for ELs and non-ELs in forecasting end-of-year calculation and WP outcomes. Differential predictive validity was associated with students' language status, depending on the type of mathematics outcome. In non-ELs, DA predicted calculation but not WP outcome, while controlling for domain-general and mathematics predictors; by contrast, in ELs, DA predicted both outcomes.

One factor to consider when designing DA for ELs is the language of DA. Orosco, Swanson, O'Connor, and Lussier (2013) developed a DA that provided linguistic scaffolds while teaching math comprehension strategies in WPs to ELs. In their DA, the language of WPs was simplified (reduced sentence lengths, removal of irrelevant information). ELs with significant reading and math difficulties also received linguistic support in three steps, which included pre-teaching of math ideas and concepts (step 1), explicit comprehension strategy instruction (step 2); and guided practice through collaborative learning with a teacher (step 3). With Step 3, students received probe questions reminding them of the strategies taught in previous steps. In a single-subject study, ELs' WP performance improved with linguistic support compared to the baseline performance, suggesting that oral language difficulties may overshadow ELs' capacity to learn from English DA instruction. Thus, delivering DA instruction in ELs' first language may provide valid results regarding ELs' mathematics learning potential.

Language of Assessment

Language is the medium through which test content is communicated to students. This is true for mathematics tests generally. On WP tests, language comprehension is centrally involved in the construct (Fuchs et al., 2019; Fuchs, Fuchs, Compton, Hamlett, & Wang, 2015). WP tests rely heavily on students' language abilities because they require students to understand the problem situation and identify semantic relations among numerals presented in text (e.g., locate the missing information, discriminate relevant information from irrelevant information). Thus, items with heavy language load may differentially function for students with the same level of mathematics ability, due to differences in their language proficiency, compromising the validity of mathematics test score interpretation for ELs. When ELs are tested in a language in which they have yet to develop proficiency, valid and fair interpretation of the test score is difficult.

Prior research demonstrates the important role of language skill in young ELs' mathematics performance within-language for both numeracy and applied problems, but cross-language transfer was not evident in applied problems (Méndez, Hammer, Lopez, & Blair, 2019; Foster, Anthony, Zucker, & Branum-Martin, 2019). Thus, ELs whose English language is low may experience greater difficulty with WPs than do non-ELs or ELs with higher English proficiency even when mathematics competence is accounted for (Lord, Abedi, & Poosuthasee, 2000; Martiniello, 2009; Shaftel, Belton-Kocher, Clasnapp, & Poggio, 2006). This is especially the case when problems pose greater linguistic demands (more prepositions, pronouns, difficult vocabulary).

Determining the language for EL testing is a complex issue that several factors, such as students' proficiency in their first and second languages (L1 and L2), and formal schooling experience such as the language of instruction, need to be considered (Kopriva, 2008). A general recommendation for preventing invalid test interpretation is to avoid testing ELs in English until their English proficiency is sufficiently developed (Katz, Low, Stack, & Tsang, 2004). Testing in L1 (which is Spanish in Spanish-speaking ELs) is therefore considered a test accommodation option for reducing the influence of factors *irrelevant* to ELs' mathematics competence on test performance (e.g., Rivera et al., 2006; Willner et al., 2008), thereby improving the test performance of students who are affected by skills irrelevant to the target construct being measured. This phenomenon is referred to as differential boost (Fuchs & Fuchs, 2001; Stretch & Osborne, 2005). A differential boost is demonstrated when an accommodation improves the test performance of ELs more than it does for non-ELs. Thus, it is expected that ELs perform better when assessed in L1 because linguistic burden on the mathematics test decreases. For example, kindergarten and first-grade Spanish-speaking ELs performed better on a mathematics test when assessed in Spanish instead of English, especially so for students with low socioeconomic status (Robinson, 2010).

At the same time, testing ELs in L1 may not represent a better option than English testing for ELs who receive instruction in English. ELs who learn content knowledge in English-only programs may develop stronger academic language (content area-specific vocabulary) and test register (the language of the test) in English than in L1 (Butler & Stevens, 1997; Solano-Flores, 2008). This is because memory processes and word representations in bilinguals are impacted by the language in which the original learning experience is encoded (Altarriba, 2003; Marian & Neisser, 2000). In fact, ELs who receive instruction in English-only programs perform better when the test is administered in English than in L1 (Abedi, Lord, & Hofstetter, 1998; Kujawa et al., 2001).

Another factor to consider when determining whether ELs should be tested in their L1 or English is the test's purpose. Whereas previous discussions have focused mainly on assessments of students' achievement level and accountability (e.g., Abedi, 2004), schools frequently use screening assessments to index risk for later LDs to allocate prevention services. If the purpose is to predict later achievement, one should consider the match between the language of the screener and the criterion assessment. L1 assessment may be less *predictive* than English testing if the goal is to index risk for academic success in English-only settings. In fact, a recommendation for screening ELs for later English reading problems has been to use the same measures and approaches as used with non-ELs (Gersten et al., 2007). Yet, there is no practice guide for screening ELs for mathematics difficulties.

It is also important to note that the decision regarding test language should be made on an individual basis (Kopriva et al., 2007). The term *English learner* fails to capture the full spectrum of language development in ELs, a heterogeneous group who vary considerably in various factors including language dominance, English proficiency, and L1 across various modalities (e.g., August & Hakuta, 1997; Ford et al., 2013; Solano-Florez & Trumbull, 2008). Recognizing ELs as emergent bilinguals (Garcia, Kleifgen, & Falchi, 2008), they develop their

L1 and English simultaneously but at different rates based on language learning contexts and history, often resulting in one language used more dominantly than another (Wei, 2000). The term *language dominance* characterizes bilingual children's relative preference or facility in one language over the other (Gathercole & Thomas, 2009; Silva-Corvalán & Treffers-Daller, 2016). Thus, beyond English proficiency, which is often defined relative to vocabulary size, language dominance should be taken into account.

Present Study

To better understand which language should DA be administered for ELs, we addressed three research questions focusing on Spanish-speaking ELs. First, do ELs perform better on DA when tested in Spanish (L1) than in English? Second, does DA's predictive validity for explaining year-end mathematics outcomes vary as a function of DA language or type of mathematics outcome? Third, does ELs' language dominance moderate the effect of DA language on its predictive validity?

We conceptualize DA as a supplementary screener to static assessments predicting two important mathematics outcomes: calculations and WPs. Calculations, particularly whole-number addition and subtraction, are foundational for advanced mathematics (Fuchs, Compton, Powell et al., 2012) and a focus of the first-grade curriculum. WPs require students to use linguistic information to construct a problem model, identify a number sentence to represent that model, and perform calculations to solve for unknown quantity. Whereas calculations serve as a platform for more complex mathematics skills and ideas, WPs are the most important school-age predictor of wages and employment in adulthood (Every Child a Chance Trust, 2009) and are emphasized in most of the strands in math curriculum at every grade. Moreover, WP solving relies more heavily on oral language comprehension than does calculation skill because of its inherent demand for students to process the text describing a problem situation while identifying semantic relations among the quantities (Fuchs et al., 2016, 2018; Swanson & Beebe-Frankenberger, 2004).

Consistent with prior studies in this line of studies (Seethaler et al., 2012; Authors, 2016), we compared DA's predictive validity against the two types of competing static assessments. This first was domain-general predictors that constitute traditional intelligence tests. Language, often indexed via oral vocabulary, is crucial in mathematics sense-making and WP solving (Jordan, Glutting, & Ramineni, 2010), particularly for ELs (Vukovic & Lesaux, 2013). Nonverbal reasoning is important in supporting various forms of mathematics development (Seethaler, Fuchs, Star, & Bryant, 2011). Also, we included two static assessments of numerical competence: a brief static screener and an extended mathematics test. A combination of these predictors creates a rigorous evaluation of DA's predictive validity.

Method

Participants

We recruited first-grade students from 75 classrooms in 17 Title-1 schools in a southeastern metropolitan public school district. Parents of 392 students provided consent (i.e., a 74.4% consent rate) via consent documents provided in English and Spanish; a Spanish-speaking staff member was available to answer parents' questions. Students were identified as ELs by their schools if they were designated as limited English proficient (LEP) and qualified to receive English as a Second Language services from the school district based on their scores on the English Language Development Assessment (ELDA; Council of Chief State School Officers, 2009) or the Tennessee English Placement Assessment (Tennessee Department of Education, 2009). A detailed description of the ELDA and TELA is provided in the online supplemental materials. Students were excluded if their first language was not Spanish (*n* = 2), if they participated in other intervention research studies (*n* = 4), or they were identified as non-EL (*n* =1). Additionally, three students moved out of district prior to pretesting, resulting in a total of 382 students who were randomly assigned to English (ENG-DA; *n* = 192) or Spanish (SPAN-DA; *n* = 190) conditions. Fourteen students moved prior to completing spring testing (*n*s = 5 and 9 from ENG-DA and SPAN-DA conditions, respectively). Attrition did not differ by condition ($\chi^2 = 1.23$, $df = 1$, $p = .28$), and movers were comparable to stayers on the incoming mathematics performance (F (1, 380) = .58, $p = .45$) and demographic variables (χ^2s < 2.64, *p*s > .10). Data for the 368 remaining students were complete. See Table 1 for demographic information by language administration condition. There were no significant differences between conditions on measures used to identify LEP status or on demographic variables, except for gender ($\chi^2 = 5.31$, $df = 1$, $p = .02$).

Measures

Incoming mathematics performance. The Number Sets Test (NST; Geary, Bailey, & Hoard, 2009) was used to measures incoming mathematics performance (see online supplemental materials for detail). It assesses the ability to quickly and accurately process quantities depicted by Arabic numerals and by sets of objects. Internal consistency reliabilities range from .70 to .90 (Geary et al., 2009).

Domain-general predictors. We assessed language (L2 expressive vocabulary and verbal knowledge) with the Wechsler Abbreviated Scale of Intelligence (WASI) Vocabulary test, which measures students' receptive (items 1-4) and expressive vocabulary. Nonverbal reasoning was assessed with the WASI Matrix Reasoning test, which indexes pattern completion, classification, analogy, and serial reasoning. Split-half reliabilities reported in the manual are above .86 (Zhu, 1999).

Mathematics predictors. We assessed numerical competency with Quantity Discrimination (QD; Chard et al., 2005; Lembke & Foegen, 2009) and Test of Early Mathematics Achievement (3rd ed.) (TEMA; Ginsburg & Baroody, 2003). QD assesses the accuracy and efficiency of making magnitude comparisons between pairs of Arabic numerals in 1 min. Test-retest reliability is .85-.99 (Clarke, Baker, Smolkowski, & Chard, 2008). The TEMA assesses informal and formal mathematics knowledge. TEMA is individually administered and takes approximately 45 min (alpha =.77).

Dynamic assessment. The Balancing Equations Dynamic Assessment (DA; Seethaler & Fuchs, 2010a) measures the degree of scaffolding required to learn unfamiliar mathematics content, specifically solving for missing variables in nonstandard addition equations. This task was selected for the following reason. Balancing equations, while critical for higher-level mathematics, is difficult for many students because they misinterpret the equal sign as an operational rather than a relational symbol (Sherman & Bisanz, 2009). Thus, balancing equations presents a learning opportunity for students and a measurement opportunity to index the amount of support a student needs to learn. For a detailed description of the DA's instruction, see Table 2 and the online supplemental materials (also see Authors, 2016).

The DA was translated to Spanish (SPAN-DA) by members of the project's bilingual (English-Spanish) research team. All graduate-student members of the research team were fluent in both Spanish and English, as a result of having been raised in homes with both languages or because they were certified, secondary-school Spanish-language teachers. The original, ENG-DA was copied electronically in its entirety to a second file. Then, one bilingual research assistant translated the entire assessment, including all scripted instructional scaffolding, worked examples, and mastery test items, to Spanish. A second bilingual research assistant then back-translated the SPAN-DA into English. We compared the back-translated ENG-DA to the original measure, noting discrepancies and resolving them.

Language dominance. Language dominance was assessed with a single-item survey asking teachers to describe the dominant language their EL students use at home and in school at the start of the school year. Based on informal classroom observations and district's home language survey, teachers answered if each student (a) speaks Spanish exclusively; (b) speaks mostly Spanish but also speaks some English; (c) speaks both Spanish and English with equal ease; or (d) speaks mostly English but also speaks some Spanish. Students were classified as Spanish-dominant ELs if they were identified by teachers as speaking exclusively or mostly Spanish (a and b); they were considered English-dominant ELs if they were either fluent in both languages (c) or speaks mostly English (d). English-dominant ELs performed significantly better on either ELDA (g = .53; F (1, 184) = 13.14, p < .05) or TELPA (g = .30; F (1, 180) = 4.24, p < .05) compared to Spanish-dominant ELs.

Year-end mathematics outcomes. We assessed calculation with the Wide Range Achievement Test (3rd ed.) - Arithmetic (WRAT-3; Wilkinson, 1993). Students answer 15 items presented orally by the tester. Items include counting objects, identifying and comparing numbers, and simple word problems. Students then have 10 min to solve 40 calculation items of increasing difficulty printed on paper. Split-half reliability reported in the manual is .94. We assessed WPS performance with Story Problems (Jordan & Hanich, 2000), which comprises 14 single-step addition and subtraction word problems of the types most often encountered in the primary grades: compare, combine, and change (alpha =.77). The tester reads aloud each item and provides one additional reading if requested.

Classroom instruction

First-grade teachers from the same school district in which the DA study took place (but collected on a different sample) completed a questionnaire in the spring, describing their whole-class math instruction. They reported an average of 233.05 min per week (SD = 75.73; range: 80 – 480) allocated to math instruction. During a typical lesson, they spent 16.71% of the lesson reviewing (SD = 5.55), 26.91% leading instruction on new content (SD = 8.79), 24.76% guiding practice (SD = 8.79), 23.90% providing independent practice (SD = 8.79), and 6.71% “other” (SD = 9.26). They also assigned 10.32 min of daily homework (SD = 6.54). They relied primarily on whole-class instruction (assigning 45.49 points, SD = 21.02, of 100 to indicate importance), with other formats as follows: small-group instruction (21.30 points, SD = 7.43), individual instruction (11.85, SD = 6.44), peer tutoring and cooperative group work (20.32, SD = 7.32), and other (1.29, SD = 2.19). Further, teachers report about 28% of their mathematics instructional time was spent teaching word problems, and the rest focused on calculations and number knowledge. Classroom instruction was provided in English.

Procedure

In September, students were screened using NST, and the survey of language dominance was obtained from teachers. In October, static assessments of the predictors (QD, TEMA, WASI-vocab and reasoning) were administered in one individual testing session. In November, DA was administered. Competing predictors were administered prior to the DA so that DA instruction would not influence performance on the other measures. Testing sessions were evenly distributed between students assigned to ENG-DA and SPAN-DA. In May, mathematic outcomes (WRAT-3 and WPS) were administered in an individual testing session. Tests were administered by research assistants and the project coordinator with 100% accuracy during training and practice administrations of the measures. All testers who administered the DA in Spanish were bilingual speaking both languages fluently or were certified Spanish teachers. Testers received one week of training, consisting of demonstration followed by practice with each other. Then, testers completed fidelity checks until they reach 100% procedural fidelity. All testing sessions were audio-recorded. Fifteen percent of sessions, distributed equally across sessions, testers, and each condition, were randomly sampled to assess fidelity of administration. Scoring agreement was above 99%.

Data Analyses

Preliminary analyses. Table 2 provides correlations and descriptive statistics by the DA condition. One-way analysis of variance indicated no differences between the DA conditions on all of the variables ($0.01 \leq F$s $(1, 366) \leq 1.32$, $.25 \leq p$s $\leq .93$). We neither found univariate outliers using the Tukey's (1977) standards nor did we detect any multivariate outliers using the blocked adaptive computationally efficient outlier nominators algorithm (Billor, Hadi, & Velleman, 2000). Class-level intra-class correlations (ICCs) were between 0 and .19. School-level ICCs were below .05, except for QD and calculation (ICCs < .08). Dependencies at the classroom-level were addressed by adjusting standard errors using Type = Complex commend with MLR estimator in M*plus* 7.4.

Primary analyses. First, we tested the mean-level differences in DA performance between ENG-DA and SPAN-DA conditions, controlling for incoming math ability (i.e., NST), using multi-group path analysis by regressing DA on NST and estimating the intercept (mean) of DA in each group. Then, we included language dominance in the model to examine the effects of language-dominance on DA performance (research question 1). Second, to examine the effects of DA language on predictive validity and the moderating role of language dominance, we ran a series of multi-group path models, comparing ENG-DA and SPAN-DA (research questions 2 -

3). Initially, we only included the static predictors (Base Model). Because students were randomly assigned to DA conditions, and these static assessments were administered prior to DA, we did not expect the relation of these static assessments outcomes to differ between the conditions. This assumption was confirmed by a series of Wald tests of parameter constraints (Wald χ^2s (1) < 3, *p*s <.05); thus, we constrained all path coefficients to be equal between the two conditions. This allowed us not only to build a parsimonious model but also to directly compare the predictive validity of ENG-DA and SPAN-DA while holding the effects of static assessments invariant between the conditions. Then, we added DA to the base model to examine DA's additional predictive value beyond what could be explained by the static assessments (DA Model). Third, we created an interaction term between language dominance and DA. We included the main effect of language dominance as a binary variable (1 = Spanish-Dominant; 0 = English-Dominant) as well as the interaction term (Interaction Model). For significant interaction effects, we calculated simple slopes of DA.

Results

Mean Differences between English-DA and Spanish-DA

Results from multi-group path analyses (Table 3) indicate that the intercept did not differ between the conditions (Wald χ^2= 0.25, *df* =1, *p* =.62), controlling for the incoming mathematics. On average, ELs in SPAN-DA condition (8.37) performed comparably to those in ENG-DA condition (8.17). However, differential boost was observed: Whereas Spanish-dominant ELs performed poorly compared to English-dominant ELs on ENG-DA (B = -1.25, p < .05), they performed at the level similar to that of English-dominant ELs on SPAN-DA (B = -0.01, p = .83).

Predictive Validity of English DA versus Spanish DA

Effects of DA language. Predictive validity results are presented in Table 4. The base model provided excellent fit to data, $\chi 2(8) = 7.59$; $p = .47$; RMSEA = 0.00; CFI = 1.00; TLI=1.00; SRMR =.02. In the Base Model, reasoning ($B = 0.08$, $p < .01$), TEMA ($B = 0.19$, $p < .01$), and QD ($B = 0.07$, $p < .05$) were significant predictors of calculation. However, only language ($B = 0.05$, $p < .01$) and TEMA ($B = 0.21$, $p < .01$) were significant predictors of WPS. When DA was added, DA Model yielded excellent fit to data, $\chi 2(8) = 5.98$; $p = .95$; RMSEA = 0.00; CFI = 1.00; TLI=1.00; SRMR =.01. ENG-DA was a significant predictor of calculation ($B = 0.19$, $p < .01$) and WPS ($B = 0.21$, $p < .01$), explaining additional 5% and 8% of their respective variance. Similarly, SPAN-DA predicted calculation ($B = 0.21$, $p < .01$) as well as WPS ($B = 0.13$, $p < .01$), explaining additional 6% and 5%, respectively.

The moderating role of language dominance. Interaction Model showed excellent fit to data, $\chi 2(8) = 7.23$; $p = .95$; RMSEA = 0.00; CFI = 1.00; TLI=1.00; SRMR =.01. The main effects of ENG-DA on calculation and WPS continued to be significant, and neither language dominance nor the interaction term were significant predictors of either outcome. However, in the SPAN-DA condition, the main effect of DA was no longer predictive of calculation ($B = 0.11$, $p = .09$) whereas the interaction term was significant ($B = 0.17$, $p < .05$), explaining an additional 2% of variance to the DA Model. Simple slope of SPAN-DA on calculation was .28 ($p < .01$) for Spanish-dominant ELs and .13 ($p = .09$) for English-dominant ELs. For predicting WPS, SPAN-DA’s main effect continued to be significant, and DA’s predictive validity did not vary as a function of language dominance.

Discussion

Accurately identifying ELs at risk for mathematics LDs without inflating false positive rates has been an uphill battle, due to the challenges associated with distinguishing true mathematics difficulties when limited English proficiency is confounded with low mathematics performance. DA has been used in clinical settings to identify language disorders (e.g., Peña, Gillam, & Bedore, 2014) but has yet to be used in school settings for early LD identification. Yet, research demonstrates DA's promise for this purpose (Authors, 2016). The purpose of the present study was to explore the effects of DA language for ELs. Spanish DA potentially minimizes the effects of limited English proficiency, which is irrelevant to students' mathematics learning ability. This would increase DA's validity as a screening measure for identifying the need for early intervention. Further, recognizing the heterogeneity of the EL population, we examined whether the effects of Spanish DA vary as a function of language dominance.

Does Administering DA in L1 Improve ELs' DA Performance?

Conceptualizing Spanish DA as a type of test accommodation, we followed the logic of test accommodation literature (Kieffer, Lesaux, Rivera, & Francis, 2009). The effectiveness of accommodation is evaluated based on whether students receiving the accommodation (SPAN-DA) perform higher than students tested without accommodation (ENG-DA) because it diminishes the negative impact of test access (language) irrelevant to the target skill being assessed (mathematics learning potential). Thus, when a test accommodation is valid, we observe a differential boost such that it improves scores only for students who lack the access skill without inflating the scores of those with appropriate access skills.

We hypothesized that Spanish DA lowers the language barrier and promotes ELs' responsiveness to DA instruction. Contrary to the findings of Orosco et al. (2013), where ELs performance on WPs improved with linguistic support, we did not identify such benefits of testing in Spanish. On average, ELs' DA scores were comparable when the testing was completed in either language. This inconsistency may be due to one of two possibilities. First, studies examining test accommodations for ELs are mostly conducted in standardized assessment settings (state or national achievement test) where students perform independently (see Kieffer et al., 2009). By contrast, students receive ample amount of scaffolds using child-friendly language in DA. In particular, we used a visual representation (i.e., number line) with which ELs could make meaning of equation solving without relying on linguistic information. Nonlinguistic schematic representations in mathematics test items have shown to attenuate negative bias against ELs (Martiniello, 2009). Thus, our DA instruction may have already been sufficiently accessible to most ELs, making testing in L1 unnecessary for most of our participants. Similarly, language load in balancing equation task is low compared to WP solving, a target task used in Orosco et al. (2013). Second, because ELs in our study received school mathematics instruction in English, DA instruction in Spanish may not have facilitated mathematical learning. We tested this last possible explanation.

Despite not finding the overall positive impact of Spanish over English testing on DA performance, we did find that the effects of DA language were conditional on ELs' language dominance. That is, we found evidence of differential boost. There was a benefit of Spanish-DA in Spanish-dominant ELs. Under the English-DA condition, Spanish-dominant ELs performed worse than English-dominant ELs, whereas they performed similarly to English-dominant ELs in the Spanish-DA condition. This finding suggests that DA instruction in English begets barriers for Spanish-dominant ELs, even when they receive classroom instruction in English. The hampered DA performance of Spanish-dominant ELs in English-DA renders the use of DA scores invalid. Interestingly (but as expected), Spanish-dominant ELs' DA performance was commensurate with that of English-dominant ELs under the Spanish-DA condition.

Results thus suggest that, for Spanish-dominant ELs, Spanish DA provides more equitable access to DA instruction than English DA. Our results confirm the importance of considering varying levels of bilingual development among ELs, as manifested in language dominance when evaluating the effect of DA language. This finding is in line with the idea that a "one-size-fits-all" approach of test accommodation for ELs does not work. Accommodation decisions should be formulated at the individual level and carefully tailored to the needs of individual students (Kopriva et al., 2007).

Does Administering DA in L1 Improve Predictive Validity?

The present study replicated results from Authors (2016) by documenting the relative contribution of DA to predicting year-end mathematics outcomes in first-grade ELs. It is important to note that both versions of the DA were significant predictors of mathematics outcomes, explaining 5% to 8% of the variance, in the presence of the competing mathematics and domain-general predictors. Although DA's unique explanatory power may not seem large, its importance as a supplementary screener should not be discounted, because it was competed against students' beginning-of-the-year mathematics ability measured with QD and TEMA, each of which has documented strong validity for predicting future mathematics outcomes (e.g., Seethaler & Fuchs, 2010b; Seethaler, Fuchs, Star, & Bryant, 2011; Seethaler, Fuchs, Fuchs, & Compton, 2011). It is important to note that DA's predictive validity was also higher than that of language ability, a well-established predictor of WP solving (e.g., Fuchs et al., 2015)

Overall, DA demonstrated similar levels of added predictive validity for explaining both calculation and WP solving, regardless of the language of administration. This finding is expected given that we did not find that Spanish-DA reveals mathematics learning potential better than English-DA on average. We note, however, that the effects of DA language may vary depending on the mathematics outcome predicted. In our previous study, English-DA showed higher predictive validity for WP solving than for calculations. In the present study, English-DA showed a similar pattern: DA explained a larger amount of variance in WPs (8%) than calculations (5%). Even though balancing-equation skill taught in DA is directly linked to calculations but not WPs, English DA showed higher predictive validity for WPs (distal outcome relative to the DA tasks) than calculation (a more proximal outcome to the DA tasks) perhaps because the mathematical learning and reasoning skill involved in DA is critical in WPs.

At the same time, we did not find a similar pattern with Spanish-DA, where predictive validity was similar for calculations and WPs. This may be because of the misalignment between the language of DA and WPs and suggests that Spanish DA's predictive validity for English WP solving is not as high as that of English-DA. This is consistent with findings from reading studies demonstrating the superiority of English over Spanish skills in accounting for English reading comprehension outcomes in Spanish-speaking ELs (Gottardo & Mueller, 2009; Mancilla-Martinez & Lesaux, 2010). Thus, when one's goal is to screen for later WP difficulties, it is preferable to select DA language that matches the language of the WP outcome measure.

Moderating Role of Language Dominance

Translating English-DA to ELs' L1 (Spanish) was intended to minimize the effect of English proficiency irrelevant to students' mathematics learning potential. However, because ELs possess unique linguistic profiles along the continuum of bilingual development, we expected the effects of DA language on predictive validity to vary as a function of students' language dominance. We found moderating effects of ELs' language dominance on DA's predictive validity, which also depends on DA language as well as the type of outcome predicted. For predicting calculations, Spanish-DA was no longer universally predictive of calculation outcome when language dominance was taken into account. Instead, Spanish-DA demonstrated stronger validity than English-DA for Spanish-dominant ELs. For English-dominant ELs, English-DA worked better than Spanish-DA. For WP prediction, language dominance did not moderate predictive validity of Spanish-DA.

Even so, there was a slight improvement in English-DA's predictive validity compared to Spanish-DA for English-dominant ELs. This opposite pattern of findings underscores the importance of simultaneously considering various factors, including language dominance and language of instruction, when selecting the language of test administration for ELs. Complex individualized decision making based on student characteristics and the nature of outcome criterion is thus critical even with DAs.

We also acknowledge that language dominance only considers relative intra-individual strength in one language over the other and does not necessarily speak to the absolute level of English proficiency (an important consideration in selecting testing accommodations). For example, Spanish-dominant ELs may have sufficiently developed English; similarly, English-dominant ELs may also have low English proficiency. In general, Spanish-dominant ELs had lower L2 vocabulary scores (M = 8.74, SD = 5.69) than English-dominant ELs (M = 13.60, SD = 5.47; F (1,366) = 69.98, $p < .05$).

To further explore complexities in determining the choice of DA language, we conducted supplementary analyses using language (L2 vocabulary), one of the competing predictors, as a moderator (see Supplementary Material for detailed method and results). These supplementary analysis results corroborate and strengthen our main findings. For calculations, a skill that does not rely as much on students' language, Spanish-DA had higher predictive validity when ELs' L2 vocabulary was lower. Such an advantage was not evident when predicting WPs. By contrast, English-DA showed stronger predictive validity when ELs' L2 vocabulary was higher with respect to predicting WPs.

Limitations

When interpreting these results, readers should consider several limitations. First, we had no direct common measure of students' Spanish proficiency, because the school district did not administer the same English proficiency test to all students. Second, to index ELs' oral language, we used expressive rather than receptive language. Although this is often the case in prior research on young ELs' mathematics (e.g., Gjicali, Astuto, Lipnevich, 2019; Foster, Anthony, Zucker, & Braunm-Martin, 2019; Hernandez et al., 2019), receptive language is more essential for the sense-making needed to profit from classroom math instruction and understand mathematics tests. Third, in the present study, language dominance was classified by teachers based on a single question, even though determining language dominance requires documenting language history and considering proficiency in both languages (Bedore et al., 2012). Fourth, DA was not administered with school-determined testing accommodations. This was necessary to permit random assignment to English versus Spanish DA conditions and to avoid the error associated with schools' test accommodations decisions (Abedi, Lord, Hofstetter, & Baker, 2000; Fuchs, Fuchs, Eaton, Hamlett, Binkely, & Crouch, 2000; Fuchs, Fuchs, Eaton, Hamlett, & Karns, 2000). Finally, information on instructional context for our study participants was derived from another sample of teachers in the same district.

Conclusion and Implications for Practice

This study's findings add to the evidentiary base supporting DA's utility in the early identification of ELs at risk for mathematics disabilities within RTI and multitiered-systems of support (Fuchs, Fuchs, & Compton, 2012). In particular, English-DA not only was effective in predicting WP solving for ELs, but also was a stronger predictor than students' language performance. This finding has important implications for practice. Because the first-grade curriculum focuses heavily on numeration and calculations, it is often difficult to forecast later WP development. DA's utility in predicting later-emerging mathematics disabilities has the potential to help schools identify children who require early interventions, especially the need for WP intervention.

Further, this study deepens understanding of how mathematics DA may be designed to address the EL population's linguistic characteristics. Spanish-dominant ELs' potential for learning calculations may be underestimated when DA is administered in English. Thus, Spanish-DA appears preferable when predicting calculation outcomes. By contrast, for predicting WP outcomes, results indicate the need to rely on English-DA for English-dominant ELs, especially when their English vocabulary is relatively strong.

References

Abedi J. (2004). The no child left behind act and English language learners: Assessment and accountability issues. *Educational Researcher*, 33, 4–14.

Abedi J., Lord C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, 14, 219–234.

Abedi J., Lord C., Hofstetter C., Baker E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational measurement: Issues and practice*, 19, 16–26.

Abedi J., Lord C., Hofstetter C. (1998, July). *Impact of selected background variables on students' NAEP math performance* [CSE Technical Report No. 478]. Center for the Study of Evaluation, University of California, Los Angeles.

Altarriba J. (2003). Does Cariño equal "liking"? A theoretical approach to conceptual nonequivalence between languages. *International Journal of Bilingualism*, 7, 305–322.

Anders Y., Rossbach H. G., Weinert S., Ebert S., Kuger S., Lehrl S., von Maurice J. (2012). Home and preschool learning environments and their relations to the development of early numeracy skills. *Early Childhood Research Quarterly*, 27, 231-244.

August D., Hakuta K. (Eds.). (1997). *Improving schooling for language-minority children: A research agenda*. National Academy Press.

Balu R., Zhu P., Doolittle F., Schiller E., Jenkins J., Gersten R. (2015, November). *Evaluation of response to intervention practices for elementary school reading* [Report NCEE 2016-4000]. National Center for Education Evaluation and Regional Assistance. https://files.eric.ed.gov/fulltext/ED560820.pdf

Bedore L. M., Peña E. D., Summers C. L., Boerger K. M., Resendiz M. D., Greene K., . . .Gillam R. B. (2012). The measure matters: Language dominance profiles across measures in Spanish-English bilingual children. *Bilingualism: Language and Cognition*, 15, 616–629.

Billor N., Hadi A. S., Velleman P. F. (2000). BACON: Blocked adaptive computationally efficient outlier nominators. *Computational Statistics & Data Analysis*, 34, 279–298.

Butler F. A., Stevens R. (1997). *Accommodation strategies for English language learners on large-scale assessments: Student characteristics and other considerations* (Vol. 448). Center for Research on Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California, Los Angeles.

Caffrey E., Fuchs D., Fuchs L. S. (2008). The predictive validity of dynamic assessment: A review. *The Journal of Special Education*, 41, 254–270.

Catts H. W., Petscher Y., Schatschneider C., Sittner Bridges M., Mendoza K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities*, 42, 163–176.

Chard D. J., Clarke B., Baker S., Otterstedt J., Braun D., Katz R. (2005). Using measures of number sense to screen for difficulties in mathematics: Preliminary findings. *Assessment for Effective Intervention*, 30, 3–14.

Cho E., Compton D. L., Fuchs D., Fuchs L. S., Bouton B. (2014). Examining the predictive validity of a dynamic assessment of decoding to forecast response to tier 2 intervention. *Journal of Learning Disabilities*, 47, 409–423.

Cho E., Compton D. L., Josol C. J. (2019). Dynamic assessment as a screening tool for early identification of reading disabilities: A latent change score approach. *Reading and Writing*. Advance online Publication.

Clarke B., Baker S., Smolkowski K., Chard D. J. (2008). An analysis of early numeracy curriculum-based measurement: Examining the role of growth in student outcomes. *Remedial and Special Education*, 29, 46–57.

Compton D. L., Fuchs D., Fuchs L. S., Bouton B., Gilbert J. K., Barquero L. A., Crouch R. C. (2010). Selecting at-risk first grade-readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology*, 102, 327–340.

Every Child a Chance Trust. (2009). *The long-term costs of numeracy difficulties*. Retrieved from http://www.everychildachancetrust.org/counts/index.cfm

Ford K. L., Cabell S. Q., Konold T. R., Invernizzi M., Gartland L. B. (2013). Diversity among Spanish-speaking English language learners: profiles of early literacy skills in kindergarten. *Reading and Writing*, 26, 889–912.

Foster M. E., Anthony J. L., Zucker T. A., Branum-Martin L. (2019). Prediction of English and Spanish kindergarten mathematics from English and Spanish cognitive and linguistic abilities in Hispanic dual language learners. *Early Childhood Research Quarterly*, 46, 213–227.

Fuchs L. S., Fuchs D. (2001). Helping teachers formulate sound test accommodation decisions for students with learning disabilities. *Learning Disabilities Research & Practice*, 16, 174–181.

Fuchs D., Fuchs L.S. (2006). Introduction to response to intervention: What, why, and how valid is it. *Reading Research Quarterly*, 41, 92–99.

Fuchs L. S., Fuchs D., Compton D. L. (2012). The early prevention of mathematics difficulty: Its power and limitations. *Journal of Learning Disabilities*, 45, 257–269.

Fuchs L. S., Fuchs D., Compton D. L., Hamlett C. L., Wang A. Y. (2015). Is word-problem solving a form of text comprehension? *Scientific Studies of Reading*, 19, 204–223.

Fuchs L. S., Fuchs D., Compton D. L., Hollenbeck K. N., Craddock C. F., Hamlett C. L. (2008). Dynamic assessment of algebraic learning in predicting third graders' development of mathematical problem solving. *Journal of Educational Psychology*, 100, 829–850.

Fuchs L. S., Fuchs D., Eaton S., Hamlett C. L., Binkley E., Crouch R. (2000). Using objective data sources to supplement teacher judgments of reading test accommodations. *Exceptional Children*, 67, 67–82.

Fuchs L. S., Fuchs D., Eaton S., Hamlett C. L., Karns K. (2000). Supplementing teachers' judgments of mathematics test accommodations with objective data sources. *School Psychology Review*, 29, 65–85.

Fuchs L. S., Compton D. L., Powell S. R., Schumacher R. F., Hamlett C. L., Vernier E., . . . Vukovic R. K. (2012). Contributions of domain-general cognitive resources and different forms of arithmetic development to pre-algebraic knowledge. *Developmental Psychology*, 48, 1315–1326.

Fuchs L. S., Geary D. C., Fuchs D., Compton D. L., Hamlett C. L. (2016). Pathways to third-grade calculation versus word-reading competence: Are they more alike or different?. *Child Development*, 87, 558–567.

Fuchs L. S., Fuchs D., Seethaler P. M., Cutting L. E., Mancilla-Martinez J. (2019). Connections Between Reading Comprehension and Word-Problem Solving via Oral Language Comprehension: Implications for Comorbid Learning Disabilities. *New directions for child and adolescent development*, 2019(165), 73–90.

García O., Kleifgen J. A., Falchi L. (2008). *From English language learners to emergent bilinguals: Equity matters—Research review no. 1*. Campaign for Educational Equity, Teachers College, Columbia University.

Gathercole V. C. M., Thomas E. M. (2009). Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Bilingualism: Language and Cognition*, 12, 213–237.

Geary D. C., Bailey D. H., Hoard M. K. (2009). Predicting mathematical achievement and mathematical learning disability with a simple screening tool: The Number Sets Test. *Journal of Psychoeducational Assessment*, 27, 265–279.

Geary D. C., Hoard M. K., Byrd-Craven J., Nugent L., Numtee C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development*, 78, 1343–1359.

Gellert A. S., Elbro C. (2018). Predicting reading disabilities using dynamic assessment of decoding before and after the onset of reading instruction: A longitudinal study from kindergarten through grade 2. *Annals of Dyslexia*, 68, 126–144.

Gersten R., Baker S. K., Shanahan T., Linan-Thompson S., Collins P., Scarcella R. (2007). *Effective literacy and English language instruction for English learners in the elementary grades*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education.

Ginsburg H., Baroody A. (2003). *Test of Early Mathematics Ability* (3rd. ed.). Pro-Ed.

Gjicali K., Astuto J., Lipnevich A. A. (2019). Relations among language comprehension, oral counting, and numeral knowledge of ethnic and racial minority young children from low-income communities. *Early Childhood Research Quarterly*, 46, 5–19.

Gottardo A., Mueller J. (2009). Are first- and second-language factors related in predicting second-language reading comprehension? A study of Spanish-speaking children acquiring English as a second language from first to second grade. *Journal of Educational Psychology*, 101, 330–344.

Grigorenko E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of learning disabilities*, 42, 111-132.

Grigorenko E. L., Sternberg R. I. (1998). Dynamic testing. *Psychological Bulletin*, 124, 75–111.

Hibel J., Jasper A. D. (2012). Delayed special education placement for learning disabilities among children of immigrants. *Social Forces*, 91, 503–530.

Jordan N. C., Glutting J., Ramineni C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learning and Individual Differences*, 20, 82–88.

Jordan N. C., Hanich L. B. (2000). Mathematical thinking in second-grade children with different forms of LD. *Journal of Learning Disabilities*, 33, 567–578.

Katz A., Low P., Stack J., Tsang S.-L. (2004). *A study of content area assessment for English language learners* [Final report prepared for Office of English Language Acquisition and Academic Achievement for Limited English Proficient Students]. ARC Associates. http://www.arcassociates.org/files/CAELLRpt9-04.pdf

Kieffer M. J., Lesaux N. K., Rivera M., Francis D. J. (2009). Accommodations for English language learners taking large-scale assessments: A meta-analysis on effectiveness and validity. *Review of Educational Research*, 79, 1168–1201.

Kopriva R. J. (2008). *Improving testing for English language learners*. Routledge.

Kopriva R. J., Emick J. E., Hipolito-Delgado C. P., Cameron C. A. (2007). Do proper accommodation assignments make a difference? Examining the impact of improved decision making on scores for English language learners. *Educational Measurement: Issues and Practice*, 26, 11–20.

Kujawa N., Cavazos G., Meyer D., Rodriguez L., Lara-Alecio R., Galloway M., Irby B. J. (2001, August). *Field-initiated research report August, 2001*. http://oaktrust.library.tamu.edu/bitstream/handle/1969.1/92873/Aldinereport%282001%29.pdf?sequence=1&isAllowed=y

Lembke E., Foegen A. (2009). Identifying early numeracy indicators for kindergarten and first-grade students. *Learning Disabilities Research & Practice*, 24, 12–20.

Limbos M. M., Geva E. (2001). Accuracy of teacher assessments of second-language students at risk for reading disability. *Journal of Learning Disabilities*, 34, 136–151.

Magnuson K. A., Meyers M. K., Ruhm C. J., Waldfogel J. (2004). Inequality in preschool education and school readiness. *American Educational Research Journal*, 41, 115–157.

Mancilla-Martinez J., Lesaux N. K. (2010). Predictors of reading comprehension for struggling readers: The case of Spanish-speaking language minority learners. *Journal of Educational Psychology*, 102, 701–711.

Marian V., Neisser U. (2000). Language-dependent recall of autobiographical memories. *Journal of Experimental Psychology: General*, 129, 361–368.

Martiniello M. (2009). Linguistic complexity, schematic representations, and differential item functioning for English language learners in math tests. *Educational Assessment*, 14, 160–179.

McFarland J., Hussar B., Wang X., Zhang J., Wang K., Rathbun A., Bullock Mann F. (2018, May). *The condition of education 2018* (NCES Report 2018-144). U.S. Department of Education, National Center for Education Statistics. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2018144

McNeil N. M., Alibali M. W. (2005). Knowledge change as a function of mathematics experience: All contexts are not created equal. *Journal of Cognition & Development*, 6, 285–306.

Méndez L. I., Hammer C. S., Lopez L. M., Blair C. (2019). Examining language and early numeracy skills in young Latino dual language learners. *Early Childhood Research Quarterly*, 46, 252–261.

Morgan P. L., Farkas G., Hillemeier M. M., Maczuga S. (2012). Are minority children disproportionately represented in early intervention and early childhood special education? *Educational Researcher*, 41, 339–351.

Morgan P. L., Farkas G., Hillemeier M. M., Mattison R., Maczuga S., Li H., Cook M. (2015). Minorities are disproportionately underrepresented in special education: Longitudinal evidence across five disability conditions. *Educational Researcher*, 44, 278–292.

Morgan P. L., Farkas G., Wu Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities*, 42, 306–321.

National Center for Education Statistics. (2017). *National Assessment of Educational progress (NAEP) mathematics report card*. https://www.nationsreportcard.gov/math_2017?grade=4

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. U.S. Department of Education. https://www2.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf

Oakes J. (2005). *Keeping track*. Yale University Press.

O'Connor R. E., Jenkins J. R. (1999). Prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading*, 3, 159–197.

Orosco M. J., Swanson H. L., O'Connor R., Lussier C. (2013). The effects of dynamic strategic math on English language learners' word problem solving. *The Journal of Special Education*, 47, 96–107.

Peña E. D., Gillam R. B., Bedore L. M. (2014). Dynamic assessment of narrative ability in English accurately identifies language impairment in English language learners. *Journal of Speech, Language, and Hearing Research*, 57, 2208–2220.

Petersen D. B., Gillam R. B. (2015). Predicting reading ability for bilingual Latino children using dynamic assessment. *Journal of Learning Disabilities*, 48, 3–21.

Petersen D. B., Gragg S. L., Spencer T. D. (2018). Predicting reading problems 6 years into the future: Dynamic assessment reduces bias and increases classification accuracy. *Language, Speech, and Hearing Services in Schools*, 49, 875–888.

Rivera C., Collum E. (Eds.). (2006). *A national review of state assessment policy and practice for English language learners*. Lawrence Erlbaum.

Robinson J. P. (2010). The effects of test translation on young English learners' mathematics performance. *Educational Researcher*, 39, 582–590.

Samson J. F., Lesaux N. K. (2009). Language-minority learners in special education: Rates and predictors of identification for services. *Journal of Learning Disabilities*, 42, 148–162.

Seethaler P. M., Fuchs L. S. (2010a). *Balancing equations dynamic assessment*. Peabody College.

Seethaler P. M., Fuchs L. S. (2010b). The predictive utility of kindergarten screening for math difficulty. *Exceptional Children*, 77, 37–59.

Seethaler P. M., Fuchs L. S., Fuchs D., Compton D. L. (2012). Predicting first graders' development of calculation versus word-problem performance: The role of dynamic assessment. *Journal of Educational Psychology*, 104, 224–234.

Seethaler P. M., Fuchs L. S., Fuchs D., Compton D. L. (2016). Does the value of dynamic assessment in predicting end-of-first-grade mathematics performance differ as a function of English language proficiency? *The Elementary school journal*, 117, 171–191.

Shaftel J., Belton-Kocher E., Glasnapp D., Poggio J. (2006). The impact of language characteristics in mathematics test items on the performance of English Language Learners and students with disabilities. *Educational Assessment*, 11, 105–126.

Sherman J., Bisanz J. (2009). Equivalence in symbolic and nonsymbolic contexts: Benefits of solving problems with manipulatives. *Journal of Educational Psychology*, 101, 88–100.

Silva-Corvalán C., Treffers-Daller J. (Eds.). (2016). *Operationalising and measuring language dominance*. Cambridge University Press.

Solano-Flores G. (2008). Who is given tests in what language by whom, when, and where? The need for probabilistic views of language in the testing of English language learners. *Educational Researcher*, 37, 189–199.

Solano-Flores G., Trumbull E. (2008). In what language should English language learners be tested. In Kopriva R. J. (Ed.), *Improving testing for English language learners* (pp. 169–200). Routledge.

Stevenson C. E., Bergwerff C. E., Heiser W. J., Resing W. C. M. (2014). Working memory and dynamic measures of analogical reasoning as predictors of children's math and reading achievement. *Infant and Child Development*, 23, 51–66.

Swanson H. L., Beebe-Frankenberger M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology*, 96, 471–491. Crossref

Swanson H. L., Howard C. B. (2005). Children with reading disabilities: Does dynamic assessment help in the classification? *Learning Disability Quarterly*, 28, 17–34.

Tennessee Department of Education. (2009). *Tennessee English language placement assessment*.

Tukey J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

Vukovic R. K., Lesaux N. K. (2013). The language of mathematics: Investigating the ways language counts for children's mathematical development. *Journal of Experimental Child Psychology*, 115, 227–244.

Vygotsky L. S. (1962). *Thought and language*. MIT Press. (Original work published 1934)

Wagner R. K., Compton D. L. (2011). Dynamic assessment and its implications for RTI models. *Journal of learning disabilities*, 44, 311-312.

Wagner R. K., Francis D. J., Morris R. D. (2005). Identifying English language learners with learning disabilities: Key challenges and possible approaches. *Learning Disabilities Research & Practice*, 20, 6–15.

Wei L. (2000). Dimensions of bilingualism. In Wei L. (Ed.), *The bilingualism reader* (pp. 3–25). Routledge.

Wilkinson G. S. (1993). *Wide Range Achievement Test 3*. Wide Range.

Willner L. S., Rivera C., Acosta B. D. (2008, October). *Descriptive study of state assessment policies for accommodating English language learners*. Center for Equity and Excellence in Education, George Washington University. https://files.eric.ed.gov/fulltext/ED539753.pdf

Zehler A. M., Fleischman H. L., Hopstock P. J., Stephenson T. G., Pendzick M. L., Sapru S. (2003). *Descriptive study of services to LEP students and LEP students with disabilities*. Development Associates.

Zhu J. (1999). *WASI manual*. Psychological Corporation.

Table 1
Student Demographics by Condition

Variable	English-DA (n = 187) n	%	Spanish-DA (n = 181) n	%	$\chi^2(df)$	p
Gender					5.31 (1)	0.02
Male	85	45.45	104	57.46		
Female	102	54.55	77	42.54		
Subsidized Lunch					2.68 (1)	0.10
No	7	3.74	2	1.10		
Yes	180	96.26	179	98.90		
Special Education Services					1.91(1)	0.17
No	182	97.33	171	94.48		
Yes	5	2.67	10	5.52		
Repeated Kindergarten					0.48 (1)	0.49
No	183	97.86	175	96.69		
Yes	4	2.14	6	3.31		
Years Receiving ELL Services						0.99
0	2	1.07	2	1.10		
1	9	4.81	13	7.18		
2	154	82.35	154	85.08		
3	18	9.63	20	11.05		
Language Dominance					.68 (3)	0.88
Speak Spanish exclusively at home and in school	13	6.95	12.0	6.67		
Speak mostly Spanish but also speaks some English	73	39.04	77.0	42.78		
Speak both Spanish and English with equal ease	94	50.27	84.0	46.67		
Speak mostly English, but also speaks some Spanish	7	3.74	8.0	4.44		
	n	M (SD)	n	M (SD)	F	p
Age (in years)	187	6.55 (0.38)	181	6.51 (0.33)	1.03 (1,366)	.31
ELDA	97	2.41 (0.81)	89	2.28 (0.78)	1.26 (1,184)	.26
TELPA	90	1.91 (0.80)	92	1.83 (0.72)	0.57 (1,180)	.45

Note. ELL = English Language Learner; ELDA = English Language Development Assessment; TELPA = Tennessee English Placement Assessment; Students receiving special education services were identified as having either one or the combinations of the following: learning disabilities, speech, and language.

Table 2

Description of DA instruction

	Type A Drawing missing circles to match the amount indicated by an Arabic numeral	Type B Solve missing number using 1 as addend and sums less than 10	Type C Solve missing number that do not use 1 as an addend with sums less than 10	Type D Sums on both sides of an equal sign and solve for a missing number
Example	6 = O O O O O _____ = 4	8 + __ = 9 or 1 + __ = 5	__ + 3 = 5	4 + 4 = 5 + __ or 3 + 6 = __ + 7
Instruction	The levels of instruction have two unsolved teaching items with which the tester models and explains a problem-solving strategy.			
Level 1	Level 1 begins with the tester presenting an already solved equation while pointing out and defining relevant mathematical terms (e.g., *equal* means *the same as*; a *plus* sign means to *add more*). The tester then suggests playing a "Hiding Game" in which one of the known quantities from the equation is covered up with a small opaque square of paper, affixed with a reusable adhesive. The tester prompts the student to solve for the "hiding number" under the paper. After the student responds, the tester affirms the response or provides the correct answer; the student removes the paper to see the missing number. Then, two unsolved items are presented in which the student is prompted to name the hiding number (i.e., the amount missing as indicated by a blank line in the equation).			
Level 2	In level 2, students receive instruction in conjunction with a 5.5-in. number line printed on paper. The number line comprises 10 half-inch squared boxes connected in a row; the boxes contain the numerals 1 to 10. Students are taught to move their finger to count the boxes on the number line while solving equations. This is designed to support understanding of the inverse relation between addition and subtraction (e.g., for 1 + __ = 3, students put their finger on the 1 on the number line and count up 2 more boxes to get to 3, revealing that 3 – 1 = 2)			
Level 3	The third level of instructional scaffolding increases support to successfully apply the number line strategy for building understanding of the inverse relation of addition and subtraction. Toward that end, different colored markers on the number line represent different parts of the equation.			

Table 3

Correlations and Descriptive Statistics for English DA (n = 187) and Spanish DA (n = 181) Conditions

	English-DA							Spanish-DA						
	L	R	QD	T	DA	CA	WPS	L	R	QD	T	DA	CA	WPS
Predictors														
Language (L)	--							--						
Reasoning (R)	.24	--						.24	--					
Quantity Discrimination (QD)	.40	.11	--					.40	.31	--				
TEMA (T)	.48	.26	.70	--				.42	.42	.68	--			
Dynamic Assessment (DA)	.28	.28	.52	.64	--			.40	.38	.52	.63	--		
Outcomes														
Calculations (CA)	.34	.31	.61	.70	.64	--		.40	.36	.54	.66	.64	--	
Word Problem Solving (WPS)	.36	.24	.45	.62	.60	.62	--	.38	.38	.38	.61	.58	.59	--
Descriptive Statistics														
M	11.30 (32.10)	7.79 (47.24)	22.67	32.34 (93.24)	8.37	17.88 (97.49)	4.52	11.30 (32.07)	7.30 (46.60)	22.08	31.80 (92.50)	8.20	17.70 (97.70)	4.40
SD	6.20 (8.97)	4.07 (7.67)	9.41	7.80 (11.77)	4.93	3.02 (15.42)	2.97	6.00 (8.68)	4.10 (7.41)	9.50	7.90 (11.00)	4.90	3.30 (15.20)	3.00
Min	0.00	0.00	0.00	10.00	0.00	8.00	0.00	1.00	1.00	0.00	3.00	0.00	9.00	0.00
Max	30.00	23.00	47.00	49.00	16.00	25.00	13.00	25.00	21.00	42.00	48.00	16.00	25.00	13.00
Skewness	0.10	1.00	-0.38	-0.37	-0.06	-0.69	0.69	0.10	0.96	-0.22	-0.62	-0.15	-0.77	0.62
Kurtosis	2.70	4.15	2.85	3.00	1.57	3.25	2.91	2.49	3.50	2.36	3.04	1.64	3.05	2.87

Note. All correlations significant at $p < .05$ except for QD and R in English DA group.
Values in the parenthesis are standardized scores
Language is Wechsler Abbreviated Scale of Intelligence (WASI) Vocabulary; Reasoning is WASI Matrix Reasoning; QD is Quantity Discrimination; TEMA is Test of Early Mathematics Ability, 3rd ed. Standardized scores for WASI Vocabulary and Matrix Reasoning are T scores (M = 50; SD = 10) and for Calculations (Wide-Range Achievement Test-Arithmetic) is standard score (M = 100; SD =15).

Table 4

Effects of DA Language on DA Performance and Effects of Language Dominance

	Base Model						Language-Dominance Model					
	English-DA			Spanish-DA			English-DA			Spanish-DA		
	B	*SE*	*p*	*B*	*SE*	*p*	*B*	*SE*	*p*	*B*	*SE*	*p*
DA												
Intercept	8.37	0.20	<.001	8.17	0.27	<.001	8.95	0.44	<.001	8.26	0.41	< .001
Number Sets Test	2.39	0.20	<.001	2.38	0.23	<.001	2.31	0.20	<.001	2.37	0.23	< .001
Language Dominance							-1.25	0.64	.049	-0.12	0.56	.827

Table 5

Results of the Multi-Group Path Analyses

		Base Model			DA Model						Interaction Model					
		Both Groups			English-DA			Spanish-DA			English-DA			Spanish-DA		
Outcome	Predictors	*B*	*SE*	*p*	*B*	*SE*	*p*	*B*	*SE*	*p*	*B*	*SE*	*p*	*B*	*SE*	*p*
Calculation																
	Language	0.02	0.02	.288	0.02	0.02	.390	0.02	0.02	.390	0.02	0.02	.349	0.02	0.02	.349
	Reasoning	0.08	0.03	.003	0.05	0.03	.062	0.05	0.03	.062	0.05	0.03	.068	0.05	0.03	.068
	QD	0.07	0.02	<.001	0.05	0.02	<.001	0.05	0.02	<.001	0.05	0.02	<.001	0.05	0.02	<.001
	TEMA	0.19	0.03	.002	0.14	0.03	.014	0.14	0.03	.014	0.14	0.03	.016	0.14	0.03	.016
	DA				0.19	0.03	<.001	0.21	0.02	<.001	0.16	0.04	<.001	0.11	0.07	.085
	Language dominance										0.05	0.32	.865	-0.15	0.36	.675
	DA x Language dominance										0.06	0.07	.354	0.17	0.07	.017
WP Solving																
	Language	0.05	0.02	.010	0.05	0.02	.012	0.05	0.02	.012	0.05	0.02	.025	0.05	0.02	.025
	Reasoning	0.06	0.04	.104	0.03	0.04	.457	0.03	0.04	.457	0.03	0.03	.468	0.03	0.03	.468
	QD	-0.01	0.02	<.001	-0.03	0.02	<.001	-0.03	0.02	<.001	-0.03	0.02	<.001	-0.03	0.02	<.001
	TEMA	0.21	0.02	.553	0.15	0.02	.143	0.15	0.02	.143	0.15	0.02	.114	0.15	0.02	.114
	DA				0.21	0.04	<.001	0.18	0.04	<.001	0.25	0.05	<.001	0.19	0.06	.001
	Language Dominance										0.14	0.37	.699	-0.28	0.29	.333
	DA x Language dominance										-0.08	0.06	.206	-0.01	0.07	.940

Note. WP = Word-Problem; DA = Dynamic Assessment; Language is Wechsler Abbreviated Scale of Intelligence (WASI) Vocabulary; Reasoning is WASI Matrix Reasoning; QD is Quantity Discrimination; TEMA is Test of Early Mathematics Ability, 3rd ed.

Coefficients of Language, Reasoning, QD, and TEMA are set to be equal across the English-DA and Spanish-DA group.