Does the Value of Dynamic Assessment in Predicting End-of-First-Grade Mathematics Performance Differ as a Function of English Language Proficiency?

Pamela M. Seethaler, Lynn S. Fuchs, and Douglas Fuchs

Vanderbilt University

and Donald L. Compton

Florida Center for Reading Research, Florida State University

Publication Date: December 2016

This research was supported in part by Grant #R324A090039 from the Institute of Education Sciences in the U.S. Department of Education to Vanderbilt University and by Awards R24HD075443 and Core Grant HD15052 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development to Vanderbilt University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Institute of Education Sciences, the U.S. Department of Education, the National Institutes of Health, or the Eunice Kennedy Shriver National Institute of Child Health and Human Development.

Abstract

The purpose of this study was to assess the added value of dynamic assessment (DA) beyond more conventional static measures for predicting individual differences in year-end 1st-grade calculation (CA) and word-problem (WP) performance, as a function of limited English proficiency (LEP) status. At the start of 1st grade, students (129 LEP; 163 non-LEP) were assessed on a brief static mathematics test, an extended static mathematics test, static tests of domain-general abilities associated with CAs and WPs (vocabulary; reasoning), and DA. Near end of 1st grade, they were assessed on CA and WP. Regression analyses indicated that the value of the predictor depends on the predicted outcome and LEP status. In predicting CAs, the extended mathematics test and DA uniquely explained variance for LEP children, with stronger predictive value for the extended mathematics test; for non-LEP children, the extended mathematics test was the only significant predictor. However, in predicting WPs, only DA and vocabulary were uniquely predictive for LEP children, with stronger value for DA; for non-LEP children, the extended mathematics test and DA were comparably uniquely predictive. Neither the brief static mathematics test nor reasoning was significant in predicting either outcome. The potential value of a gated screening process, using an extended mathematics assessment to predict CAs and using DA to predict WPs, is discussed.

Does the Value of Dynamic Assessment in Predicting End-of-First-Grade Mathematics Performance Differ as a Function of English Language Proficiency?

With dynamic assessment (DA), a tester administers a structured lesson on novel tasks while providing guidance and feedback to help the examinee learn the tasks. The resulting score indexes responsiveness to the assisted learning experience and is thought to reflect capacity to profit from future instruction. For many years, conversations (e.g. Kern, 1930; Penrose, 1934; Rey, 1934, as cited in Grigorenko & Sternberg, 1998) have focused on whether DA may serve as an alternative or supplement to conventional static assessments, in which examinees respond without assistance.

On the other hand, "*static*" tests differ from DA in an important way. With static testing, a student's performance is measured without allowing for an instructional interaction between the tester and the student. Static tests assess a student's current level of skill and knowledge, not the student's ability to acquire new skill or knowledge. The goal with DA is to reduce an important problem with static tests: They reveal only two states, unaided success or failure, which masks distinctions among children who cannot perform a task independently but succeed with varying levels of assistance (Sternberg, 1996; Tzuriel & Haywood, 1992). In a related way, the same low score earned on a static test by two students may not indicate the same potential for school learning. One low score may be due to inadequate prior learning opportunity, and although the child presently cannot perform a task, he/she profits nicely from a well-designed classroom program. The other low score, however, may reflect underlying learning challenges that result in inadequate response to the same classroom instruction.

This issue may be especially salient for children with limited English proficiency (LEP), who are more likely than non-LEP children to score poorly on static tests of learning potential due to inadequate exposure to the language demands imposed by those tests (Abedi & Lord, 2001). DA, which can be designed to minimize those demands and which directly assesses the child's capacity to learn, may represent an approach for enhancing predictions about year-end academic outcomes. Yet, in schools today, static tests are the dominant form of assessment for forecasting students' capacity to profit from general education and identifying students who require intervention.

Yet, some may wonder about the feasibility of using DA in a responsiveness-to-intervention (RTI) screening process, in which the goal is to identify students who are at-risk for academic problems early and provide them with intervention to prevent the onset of severe and intractable deficits. With RTI, screening typically occurs universally (with all students in a school) using a brief (5 to 10-min) static test. DA administration, by contrast, requires more time (typically 30-45 min). Yet, brief static tests often result in unacceptably high rates of false positives (e.g., Jenkins, Hudson, & Johnson, 2007; Seethaler & Fuchs, 2010), and each false positive error results in costly intervention, which is unnecessary. If DA reduces the rate of false positives, as has been shown for non-LEP students in mathematics (e.g., Fuchs, Fuchs, Compton, et al., 2008; Seethaler et al., 2011), it decreases expenditures on intervention for students who would respond nicely without that costly intervention.

Moreover, as noted by Fuchs, Fuchs, and Compton (2012), DA can be incorporated into a gated screening process, in which a brief static test is first administered universally, with a cut-point set to avoid false negatives. DA then is administered to the subset of students scoring below the cut-point to confirm or disconfirm the potential risk suggested in the first-round universal screen. Although such a gated screening system requires more testing time than sole reliance on a brief static test, it is less expensive by decreasing incorrect placements into intervention.

The purpose of the present study was to investigate whether DA provides added value, beyond a brief static screener, in predicting students' year-end performance, as a function of LEP status. Our focus was first-grade mathematics. To create a stringent test of our question, we also considered two other types of competing predictors: (a) a more extended, comprehensive static mathematics test and (b) domain-general static tests of learning potential, assessing individual differences on variables associated with our year-end outcomes of interest: calculations (CAs) and word problems (WPs). In this introduction, we explain our rationale for focusing on these outcomes; provide a context for DA design features, using prior work, and explain the present study's approach; describe how this DA was designed to address the needs of LEP students; and explain the rationale for our competing predictors.

Focus on CA and WP Outcomes

The present study focused on DA's value as a predictor of learning for two types of mathematics development: CA versus WP performance. These two forms of mathematics development are transparently different. Whereas CA problems are set up for solution, WPs require students to use linguistic information to construct a problem model; students must identify missing information, construct a number sentence, and set up a CA problem for solution. Beyond the transparent differences between CA and WP, prior work suggests that the cognitive characteristics underlying development in CA versus WP differ, as well (e.g., L. S. Fuchs, Fuchs, Stuebing, et al., 2008; L. S. Fuchs, Geary, Compton, Fuchs, Hamlett, Seethaler, et al., 2010; Swanson & Beebe-Frankenberger, 2004).

We selected CAs and WPs as the first-grade mathematics outcomes for the following reasons. First, whole-number addition and subtraction CA skills are foundational for more advanced mathematics (Fuchs et al., 2006; Fuchs, Powell, et al., 2012) and represent a strong focus of the first-grade curriculum. Second, although CAs are foundational to WP skill (Fuchs et al., 2006, 2011), CAs are not a sufficient pathway to WP competence (Fuchs, Fuchs, Compton, Hamlett, & Wang, 2014). Fuchs et al. (2006) used path analysis of arithmetic, arithmetic computation, and arithmetic WP performance of third-grade students to show that WP skill is linked significantly with CA skill, making CA skill necessary but not sufficient for solving WPs. Additionally, effect sizes for CA intervention on WP outcomes are substantially lower than on CA outcomes, and CA intervention narrows the CA achievement gap even as the WP achievement gap widens (Fuchs, Geary, et al., 2012). We also note that WPs are the best school-age predictor of employment and wages in adulthood (e.g., Every Child a Chance Trust, 2009), that WPs can be a persistent deficit even when CA skill is adequate (Swanson et al., 2008), and that the cognitive processes involved in WPs differ from and are more numerous than those underlying CAs (e.g., Fuchs, Geary, et al., 2010a,b; Fuchs et al., 2006; Fuchs, Fuchs, Stuebing et al., 2008). So WPs and CAs may represent distinct components of mathematics competence, and screening exclusively for CA outcomes may miss students who require intervention. For these reasons, we included both CAs and WPs as predicted outcomes.

DA Design Features and Present Study's Approach

Studies vary with respect to how DAs are structured. In terms of scoring, DAs may quantify responsiveness to the assisted learning experience as improvement from unassisted pretest to unassisted posttest (Ferrara, Brown, & Campione, 1986) or the amount of scaffolding required during the assisted learning experience to reach criterion performance (Murray et al., 2000; Spector, 1992). Interaction style is another dimension along which DAs vary. With standardized DAs (Ferrara et al., 1986), testers rely on a fixed series of prompts; other DAs (Tzuriel & Feuerstein, 1992) are individualized, with testers addressing the specific obstacles examinees reveal. A third dimension along which DAs differ is the nature of the tasks used for the assisted learning experience, which may teach domain-general cognitive abilities (Budoff, 1967; Feuerstein, 1979) or cognitive abilities presumed to underlie the academic domain to be predicted (Swanson & Howard, 2005) or domain-specific (reading or mathematics) tasks (Bransford et al., 1987; Campione, 1989; Campione & Brown, 1987; Spector, 1992).

The DA literature also varies in terms of research questions and methodological features. Researchers who index pre-posttest improvement typically investigate whether that score distinguishes between individuals with and without a pre-established diagnosis associated with poor learning (Tzuriel & Feuerstein, 1992). By contrast, researchers who index the degree of required scaffolding typically examine the value of that score in predicting a learning outcome external to the DA (Spector, 1992). This second type of study can be further categorized in terms of whether competing static predictors of outcome are considered and whether the external learning outcome is assessed concurrently with the DA or at a future time. Studies that control for competing predictors and measure the external learning outcome at a later time impose a more stringent test of DA’s value. In the present study, we examined the contribution of DA in forecasting future external learning while considering the contribution of competing predictors.

Addressing the Needs of Students with LEP using DA

Nearly one in five students in U.S. public schools speaks a language other than English at home (Wagner et al., 2005), an increase of 72% between 1992 and 2002 (Zehler et al., 2003). Of these students, Hispanic Americans represent the fastest growing group (Aud, Fox, &

KewalRamani, 2010), and three-fourths speak Spanish as their primary language (Yates & Ortiz, 2004). Increasing numbers of students with LEP in public schools present special challenges for forecasting academic failure. Students with LEP may perform poorly on assessments due to the language challenges of the test rather than to mathematical difficulties, producing error in estimates of academic difficulty (Kopriva et al., 2007).

Suggestions vary on how to increase precision. One possibility is to use test translation as an accommodation (Stansfield, 2011). Robinson (2010) found that Spanish-speaking kindergarteners and first graders scored higher on mathematics tests when tested in Spanish. Yet, native-language testing is costly, requiring translation and bilingual test personnel. Moreover, an alternative view is that students perform better on mathematics tests that match the language of instruction, rather than the language spoken at home, and that prediction of response to future instruction may be enhanced through the use of English in screening assessments. The assumption is that exposure to mathematics concepts and vocabulary may be limited to English. A meta-analysis on testing accommodations for students with LEP (Kieffer, Lesaux, Rivera, & Francis, 2012) concluded that matching the language of tests with the language of instruction produces stronger scores.

In the present study, for the DA, we took the following approach, which integrates both perspectives. We offered general test directions (pertaining to non-mathematics content) in Spanish (as well as English) for students with LEP, although very few students took the Spanish options. However, the DA's instructional scaffolding, which was conducted exclusively in English using English mathematics terminology, provided explicit instruction and other supports for English symbol and number names. At the same time, the DA was structured to minimize demands on language, so instructional prompts incorporated gesturing and relied on succinct, direct, and simple verbal explanations focused on the essence of the mathematical ideas.

Competing Predictors

We included four measures to compete with DA in predicting individual differences in CA and WP outcomes: a brief static mathematics test; a more extended, comprehensive static mathematics test; a test of reasoning ability (associated with individual differences in CA and WP skill); and a test of vocabulary knowledge (associated with individual differences in WP skill). The brief static test was Quantity Discrimination (QD; Chard et al., 2005), which measures the speed and accuracy with which children make magnitude comparisons between pairs of Arabic numerals. We included QD because it is (a) a widely used screener in RTI systems, (b) a documented predictor of future mathematics performance (Chard et al., 2005; Clarke & Shinn, 2004; Lembke & Foegen, 2005; Seethaler & Fuchs, 2010b; Seethaler et al., 2012) and (c) not redundant with the transparent content of the CA or WP outcome. Given that QD predicts CA outcomes and that WP skill partially depends on CAs, we hypothesized that QD would explain variance in both types of mathematics outcomes.

However, because QD is a speeded assessment that focuses on a narrow form of early numerical competency, we also included a non-speeded, extended mathematics test that assesses a broader range of mathematical competencies: Test of Early Mathematics Ability-3 (TEMA: Ginsberg & Baroody, 2003). TEMA assesses numbering (counting by 1, by 10s, or from a given number, identifying numbers before/after a given number), number comparison (identifying smaller/larger quantities from collections of items; selecting the Arabic numeral closer to a given numeral), CAs (solving mental, nonverbal addition problems; demonstrating addition of one or more objects with manipulatives; 2-digit addition/subtraction; speeded number combinations), understanding cardinality (counting collections of printed objects), equal partitioning (using manipulatives to show how sisters can fairly share a set of objects), numeral literacy (reading/writing numerals from 1 to 4 digits), additive commutativity principle, word problems, and base-10. TEMA predicts CA and WP skill among non-LEP first graders (Seethaler et al., 2012). We were interested in competing the predictive value of DA, also a domain-specific measure, against another domain-specific but static assessment of similar duration (~45 min). Because TEMA assesses a wide range of numerical competencies and explicitly indexes items mirroring both the CA and WP outcomes, we hypothesized it would predict both outcomes.

The other two predictors assessed domain-general cognitive and linguistic resources, vocabulary knowledge and reasoning, which represent two important components of traditional IQ (learning potential) tests and have been specifically linked to mathematics development at first grade: reasoning in the case of CAs (Fuchs et al., 2012) and vocabulary in the case of WPs (Fuchs, Geary, et al., 2010a; 2010b; Seethaler et al., 2010; Seethaler et al., 2012).

Extensions to the Literature

We identified four prior studies examining the added value of a standardized DA over competing predictors, while quantifying DA performance as the amount of required scaffolding. Speece et al.'s (l990) DA involved a domain-general task associated with overall cognitive ability: solving matrices from IQ tests. The contribution of DA over verbal IQ, pre-DA matrices performance, and language ability in accounting for variance in a concurrently administered CA test was statistically significant. Swanson and Howard (2005) centered DA on cognitive abilities presumed to underlie mathematics performance: phonological and semantic working memory. The semantic DA score explained a significant proportion of variance in concurrent CA performance, while controlling for verbal IQ and pre-DA working memory. These results suggest the potential value of DA, but neither study focused on a domain-specific (mathematics)

DA task, included extended static mathematics tests as competing predictors, or assessed mathematics outcomes at a later time.

Two more recent prior studies incorporated these methodological features. Fuchs, Fuchs, Compton, et al. (2008) designed a DA to teach novel mathematics content (pre-algebra tasks) to third graders. In the fall, students were assessed on cognitive and linguistic variables associated with WP skill, on initial CA and WP performance, and on the DA. Students were then randomly assigned to 16 weeks of validated (i.e., through five large-scale randomized control trials, published in peer-review journals) versus conventional (i.e., traditional classroom instruction) WP instruction. Near the end of the school year, students were assessed on WP outcomes that were proximal and distal to the instruction they received. The type of WP instruction students received (validated vs. conventional) adequately accounted for WP development on the measure proximal to instruction. However, language, pretreatment mathematics skill, and DA were uniquely predictive in forecasting year-end performance on the WP measure distal to instruction.

Seethaler et al. (2011) extended the 2008 study by using a DA appropriate for first graders, for which the novel task was solving nonstandard equations; by assessing DA’s predictive value at the start of first grade, when forecasting mathematics outcomes has proved particularly challenging (Compton et al., 2010; Johnson et al., 2009); and by forecasting CA as well as WP outcomes. Predictors were a brief static measure of magnitude comparison, an extended static measure of numerical competencies, domain-general measures of language and reasoning, and the DA. For forecasting year-end CA skill, the magnitude comparison test was the single most powerful predictor, followed by the extended test of numerical competencies and the DA. By contrast, for forecasting WP development, DA was the single most powerful predictor, followed by the two domain-specific mathematics tests and the language measure. Thus, different constellations of predictors appear required for predicting year-end CA versus WP skill, and DA may be especially useful for predicting WP outcomes.

An important limitation across these four prior studies, however, is that their samples were almost entirely non-LEP students. This is problematic, given the increasing numbers of students enrolled in U.S. public schools who are not native English speakers. The present study therefore extends previous work concerning the use of DA by examining whether DA offers added value, beyond competing predictors, as a function of children's LEP status. Finding added value for DA (or for other tests) beyond the brief static screener would suggest the need to consider a gated screening process. Finding that predictive relations are similar regardless of LEP status would provide support for testing students in the language of instruction (English). Finding varying predictive patterns as a function of LEP status would suggest the need for a different screening battery for children with and without LEP.

Method

Participants

We recruited first graders from 44 classrooms in 10 Title 1 schools in a southeastern metropolitan district. Parents of 600 students (i.e., 71% of students in the 44 classrooms) provided consent (documents were available in English and Spanish). We excluded 31 students for whom neither English nor Spanish was the primary language spoken at home; 18 students for whom English was not the primary language spoken at home and parents refused to answer questions about the home language (so that we could not determine if Spanish was the primary language spoken); and 28 students who did not complete the measure of incoming mathematics performance during screening. From the remaining consented students, we included all teacher-designated Spanish-speaking students with beginning or pre-functional levels of English language proficiency (i.e., 166). Students were assigned a quartile score of 1 (lowest) to 4 (highest) using results from the Number Sets Test (Geary et al., 2009; see Measures for Sample Selection); we then randomly selected students with English proficiency to match the quartile scores of all the Spanish-speaking students. This resulted in 332 students: 166 with English-language proficiency and 166, without. Cross-checking teacher designations with school records showed 24 Spanish-speaking students were misidentified by their teachers as having little English proficiency; they were, in fact, proficient and subsequently switched to the English-speaking group of students. For the final sample, Spanish-speaking students with beginning or pre-functional levels of English language proficiency, according to district measures, whom the school district designated as in need of English as a Second Language services, comprised the LEP group. (Note: The terms *English learner* or *English-language learner* and *LEP* may be used interchangeably to refer to students whose primary language is not English and who have not developed fluency in the English language. We use the term LEP in the present study because it is the official federal term used for identification and reporting purposes.) Native English speakers and English-proficient students comprised the non-LEP group.

This resulted in 142 students in the LEP group and 190 in the non-LEP group (with all schools and classrooms represented). Forty students moved prior to spring testing: 13 (9% of) LEP students and 27 (14% of) non-LEP students, for a final sample of 129 LEP and 163 non-LEP children. On the Number Sets Test (Geary et al., 2009; a measure administered at the start of the study to ensure a representative sample) and on demographic variables, the 40 students who moved were comparable to students who completed testing, with one exception: a greater proportion of students in the final sample received subsidized lunch, χ^2 (1, N = 332) = 67.76, $p < .001$. (Subsidized lunch status was not available for nine exiting students.) Data for the final sample of 292 students were complete. See Table 1 for demographics by LEP status. As shown, compared to the non-LEP group, proportionately more students in the LEP group were Hispanic and received reduced or free lunch. Otherwise, the groups were comparable.

Measures for Sample Selection

English proficiency. The school district administers the English Language Development Assessment (ELDA; Council of Chief State School Officers, 2009) each spring to students receiving ESL in grades K-12. Teachers complete observation inventories, focused on skills in four domains: listening, speaking, reading, and writing. Reports provide a performance level for each domain and for the composite score (1 = pre-functional; 5 = fully English proficient). As per the test developer, alpha at grades 1-2 for the four domains is between .93 and .96. In the LEP group, there were 10 students with a composite score of 1, 49 student with a composite score of 2, and 41 students with a composite score of 3.

To screen newly registered students (without ELDA scores from the previous spring) for ESL eligibility, trained school personnel administer the Tennessee English Placement Assessment (TELPA; Tennessee Department of Education, 2009), which is based on ELDA. TELPA measures listening, speaking, reading, and writing with 21, 8, 16, and 10 items, respectively, using multiple-choice, oral, and written response formats. Items progress in difficulty with ceiling rules guiding termination of each section. Cutpoints based on total scores for each domain yield a composite score of 1, 2, or 3, with 1 requiring at least 1 hour of ESL services per day. There were 29 students in the LEP group with a composite score of 1.

Obtaining a representative sample. To obtain a representative sample on incoming mathematics performance, we used the Number Sets Test (Geary et al., 2009), which measures the ability to quickly and accurately process quantities. Each page shows 36 pairs of squares joined in domino-like rectangles. Each side of a domino shows an Arabic numeral or set of small geometrical shapes (0 to 9). Dominos are displayed in lines of five across six rows, followed by two lines of three 3-square dominos. At the top center of each page, a target number appears: 5 on the first two pages and 9 on the last two pages. Students circle dominos that sum to the target number in 60 sec per page 5 and 90 sec per page 9. On each page, 18 items match the target number, 12 are larger, and six are smaller. Six items per page include 0 or an empty square. The tester models practice items with target numbers 4 and 3. Research assistants (RAs) worked in pairs to administer the test to large groups. One RA read aloud from the administration script and timed the test while the other circulated to monitor students. First-grade performance is consistent across target numbers/item type and can be combined to form frequency of hits (alpha, $\alpha = .88$), misses ($\alpha = .70$), correct rejections ($\alpha = .85$), and false alarms ($\alpha = .90$) (Geary et al., 2007). The score is the *z*-score for hits minus the *z*-score for false alarms.

Predictor Measures

Domain-general cognitive/linguistic predictors. With the Wechsler Abbreviated Scale of Intelligence (WASI; Psychological Corporation, 1999) - Vocabulary, the tester presents four pictures of objects for the child to name; each response is scored as 0 or 1. For remaining items, the tester says a word for the child to define; each response is scored as 0, 1, or 2 depending on quality. Testing is discontinued after five consecutive scores of 0. Split-half reliability is .86-.87 (Zhu, 1999). WASI Matrix Reasoning measures nonverbal reasoning with pattern completion, analogy, classification, and serial reasoning. Children see a matrix with one piece missing and select the missing piece from five choices. Testing is discontinued after four of five missed items. Reliability is .94 (Zhu, 1999).

Static mathematics predictors. The brief static mathematics predictor was QD (Chard et al., 2005; Lembke & Foegen, 2009), with which students have 1 min to name the larger of two numbers (ranging from 0 to 20), presented in 63 pairs of boxes. Test-retest reliability is .85-.99 (Clarke et al., 2008).

The extended static mathematics predictor was TEMA (3rd ed.; Ginsburg & Baroody, 2003), which assesses informal and formal mathematics knowledge with 72 items of increasing difficulty. Students have multiple trials for each item. Testing is discontinued after five scores of 0, and the tester ensures a basal of five correct responses is met. Alpha for 6 year olds is .95.

DA. The Balancing Equations Dynamic Assessment (Seethaler & Fuchs, 2012, revised from Seethaler et al., 2011) measures the degree of scaffolding required to learn unfamiliar mathematics content. We selected balancing equations (solving for missing variables in nonstandard equations) for the DA content for three reasons. First, the task is novel for first-grade children and therefore can be used to index the amount of scaffolding needed to learn. Second and relatedly, elementary school students often misinterpret the equal sign as an operational rather than as a relational symbol (McNeil & Alibali, 2005; Sherman & Bisanz, 2009) and thus have difficulty when the missing number is not immediately after the equal sign (i.e., in nonstandard equations). Third, because balancing equations is important for higher level mathematics, it is valuable for students to learn.

The DA includes four types of equations. Equation Type A shows a numeral on one side of the equal sign and a set circles on the other side; the number of circles is less than the numeral indicates (e.g., 6 = O O O ______ or O O ______ = 4). The child draws circles on the line to balance the equation. Items are presented with numerals on either side of the equal sign. The three remaining equation types involve addition statements, for which the child writes a missing numeral on a blank (sums <10). For Equation Type B, the missing addend is in the second position of the equation; one of the two addends is 1; the sum is always on the right side of the equal sign (e.g., 8 + __ = 9 or 1 + __ = 5). For Equation Type C, the missing variable in always in the first position of the equation; none of the addends is 1 (e.g., __ + 3 = 5). For Equation Type D, one addition statement appears to the left side of the equal sign, and another addition statement appears to the right of the equal sign. The blank is in the first or second position of the addition statements to the right of the equal sign (e.g., 4 + 4 = 5 + __ or 3 + 6 = __ + 7). Because equation types increase in difficulty, success with one equation type should promote understanding of the subsequent equation type.

Administration and scoring follows Fuchs, Fuchs, Compton, et al. (2008) and Seethaler et al. (2011). For each equation type, the tester begins by assessing mastery of that equation type. If mastery is demonstrated, the child advances to the next equation type. If not, instructional scaffolding (see below) begins with the least explicit level of scaffolding. Mastery testing then recurs. If mastery is achieved, the student progresses to the next equation type. If not, the next more explicit level of instructional scaffolding ensues, and mastery testing follows. In this way, three increasingly explicit levels of scaffolding occur. If the student fails to master a given equation type after all three scaffolding levels for that equation type, the DA is terminated. Each mastery test comprises six items of the targeted equation type.

Mastery test items are not used for scaffolding. If a student asks for help during mastery testing, the tester responds, “Just try your best.” If a student writes nothing for 15 s, the tester asks, “Can you try this one?” while pointing to the first item. If after 15 more s the child has not written anything, the tester asks, “Are you still working or are you stuck?” If the child says he/she is stuck or if 15 more s elapse with no observable attempt to solve the problem, the tester removes the mastery test and begins the first/next level of scaffolding.

Each equation type involves three levels of instructional scaffolding of increasing explicitness. Scaffolding is scripted to ensure consistency in language and procedures. Student attention is maintained with frequent questions and participation. For each equation type, the first (least explicit) scaffolding level involves the tester presenting a solved item (no missing number) while pointing out and defining mathematical terms (e.g., *equal* means *the same as*; *plus* sign means to *add more*) and then playing the Hiding Game in which one quantity in the equation is covered with a small square of paper, affixed with a reusable adhesive. In the Hiding Game, the tester prompts the child to solve for the hiding number. The tester affirms the response or provides the correct answer, and the child removes the paper to see the number. This recurs two more times. For the second scaffolding level, instruction occurs with a number line, comprising 10 squared boxes connected in a row; the boxes contain numerals 1 to 10. Students are taught to move their finger to count boxes on the number line to find the missing number that solves the equation. This is designed to support understanding of the inverse relation between addition and subtraction (e.g., for 1 + __ = 3, students put their finger on 1 and count up 2 more boxes to get to 3, revealing that 3 – 1 = 2). The third, most explicit scaffolding level increases support to successfully apply the number line strategy, using different colored markers to represent parts of the equation. Students are taught to circle the number line boxes up to the known addend on the number line with a green marker. Next, students circle additional boxes on the number line in a contrasting color until they reach the given sum in the equation. Students solve for the missing addend in the equation by counting how many additional boxes on the number line were needed to reach the given sum.

Worked examples completed during instructional scaffolding are not displayed during mastery testing. However, all materials necessary for applying the DA's scaffolding strategies are available for children to use (including during the initial mastery test). For example, number lines are available on the desk for students to use during the DA session, even though the tester does not explicitly reference the number line (except during the second and third levels of scaffolding for each equation type). Also, students are not prompted to use or penalized for their choice of a solution strategy during mastery tests. An equation type is considered mastered if the student answers at least five of the six items correctly on a mastery test, in which case the DA progresses to the next equation type.

DA scores range from 0 to 16. Zero indicates a child did not master any of the four equation types; 16 indicates mastery of all four equation types on the first administration of each mastery test (without any instructional scaffolding). The tester subtracts 1 point from the maximum of 16 for each level of scaffolding used. For example, if a child demonstrates mastery on the first administration of the mastery test for Equation Types A, B, and C (without any instructional scaffolding), but requires three levels of scaffolding to master Equation Type D, the tester subtracts 3 points from 16, for a score of 13. If a child requires three levels of scaffolding to master Equation Type A, requires two levels of instructional scaffolding to master Equation Type B, and fails to master Equation Type C (terminating the DA), the child loses 3 points for Equation Type A, 2 points for Equation Type B, and 4 points each for Equation Types C and D, for a score of 3. To index internal consistency reliability, the score from each equation type was correlated with the DA total score, using the subset of students who had not reached a ceiling prior to the administration of that equation type. For Equation Type A, $r = .61$; for Equation Type B, $r = .88$; for Equation Type C, $r = .85$; for Equation Type D, respectively, $r = .81$.

Although students can earn the same score in multiple ways (e.g., students might earn a score of four by succeeding on all four equation types after all prompts; succeeding on the first equation type without any prompts and then failing on the second equation type; or succeeding with the first two equation types after two prompts and then failing on the third equation type), we found no examples of such extreme patterns, which is in line with internal consistency reliability estimates (.61, .88, .85, and .81). Also, alternative paths to the same score are possible for each of the other predictors (as is the case for most assessments): for example, on QD, by succeeding with pairs of numerals representing small or large distances; on TEMA, vocabulary, and reasoning, by failing easier items before passing harder items (before the ceiling is reached); and on TEMA, by demonstrating competency on different profiles of mathematics domains.

Also, it is worth noting that the DA's added value may be due to its dynamic nature or its focus on balancing equations. We note, however, that this issue pertains across DA prediction studies and uses, because the DA's content is, by definition, novel – yet to be learned. So a static test of balancing equations (without instructional scaffolding) would produce a severe floor effect that accounts for little or no variance in either study outcome. It is therefore necessary that the mathematics content on DAs differ from the content on static predictors.

Mathematics outcomes. We assessed CA performance with the Wide Range Achievement Test (3rd ed.)-Arithmetic (WRAT-3; Wilkinson, 1993). Students answer 15 items presented orally: counting objects (3 items), identifying Arabic numerals (5 items), holding up a specified number of fingers (2 items), naming the larger of two numbers (2 items), and answering simple word problems (3 items). Students then have 10 min to write answers on paper to 40 CA items of increasing difficulty (grades K-12). All students finished in <10 min. As per the manual, reliability is .94.

We assessed WP performance with Story Problems (Jordan & Hanich, 2000), which includes 14 single-step addition and subtraction WPs of the types most common in the primary grades: compare, combine, and change. The tester reads aloud each problem and provides one additional reading if requested. For each problem, students have 30 s to write an answer on their paper copy. All students finished each WP within the 30-s limit. Alpha on this sample was .76.

Procedure

Testing on all predictors/outcomes occurred individually. On all measures, students were tested in English, but we offered general test directions (pertaining to non-mathematics content) in Spanish for students with LEP (although very few students took the Spanish option). All RAs were bilingual. In October and November, WASI Vocabulary, WASI Matrix Reasoning, QD, and TEMA were administered in one session and the DA was administered in a second session. Each session took less than 60 min to complete. Competing predictors were administered prior to the DA so the DA would not influence performance on the static tests. In May, the year-end mathematics outcomes were administered in a single session. Tests were administered by RAs and the project coordinator who demonstrated 100% accuracy during training and practice administrations. All sessions were audiotaped. To assess fidelity of administration, we randomly sampled 16% of tapes, stratifying by session, tester, and LEP/non-LEP status. Agreement was 98.8%. Data were independently entered twice and compared for discrepancies, which were resolved against the original protocols.

Data Analysis and Results

We used 1-way analysis of variance to test for differences between the LEP and non-LEP groups for each of the five predictors and the two outcome measures. (Note: distribution of the DA predictor and the WP outcome variables were positively skewed. Because results for the original and the transformed variables were similar, we included results of the original variables for ease of interpretation.) See Table 2 for means and standard deviations (*SD*s). There were no significant differences on reasoning, DA, CAs, or WPs. Significant differences, all favoring non-LEP students, occurred on vocabulary, $F(1, 291) = 47.68$, $p < .001$, ES = 0.70; QD, $F(1, 291) = 11.08$, $p < .001$, ES = 0.40; and TEMA $F(1, 291) = 13.35$, $p < .001$, ES = 0.40. We note that we did not expect differences on reasoning, because it does not involve language. We did not expect differences on DA, because it was designed to minimize language demands; it addressed content equally novel for LEP and non-LEP children; and it supported LEP students' facility with English terms and number names. Also, we did not expect differences on the spring (CA or WP) measures, because those tests were administered after these first-grade LEP children had been immersed in English for instruction and social interactions at school for an entire academic year, and language minority students use more English with their parent(s) across time (Mancilla-Martinez & Kieffer, 2010).

See Table 2 for correlations by LEP status and Table 3 for results of regression analyses, which were conducted separately for the LEP and non-LEP groups and separately for the CA and WP outcomes. In predicting the CA outcome, results were as follows. For LEP students, the five predictors accounted for 47.8% of the variance, $F(5, 123) = 22.51$, $p < .001$. TEMA and DA each uniquely explained variance. Follow-up analyses comparing the predictive value for TEMA versus DA, with both entered simultaneously into the model, favored the predictive value of TEMA over DA, $t(123) = 3.65$, $p < .001$. For non-LEP students, the five predictors accounted for 48.8% of the variance, $F(5, 157) = 29.95$, $p < .001$. TEMA was the only predictor to uniquely explain variance.

In predicting the WP outcome, results were as follows. For LEP students, the five predictors accounted for 39.2% of the variance, $F(5, 123) = 15.88$, $p < .001$. Vocabulary and DA each uniquely explained variance. In follow-up analyses, a significant effect favored DA over vocabulary, $t(123) = 2.85$, $p = .003$. For non-LEP students, the five predictors accounted for 47.2% of the variance, $F(5, 157) = 28.06$, $p < .001$. TEMA and DA each uniquely explained ariance. Follow-up analyses indicated no significant difference in predictive value for TEMA versus DA, $t(157) = .84$, $p = .20$.

Discussion

The purpose of this study was to investigate whether DA provides added value, beyond a brief static screener, in predicting students' year-end performance, as a function of LEP status. Our focus was first-grade mathematics. To create a stringent test of our question, we included three additional predictors: a more extended, comprehensive static mathematics assessment (compared to QD) and two domain-general static tests, assessing individual differences on variables associated in the literature with CA and WP development (reasoning and vocabulary). Before we discuss the major findings of the study, focused on DA, we briefly address QD results because in the present study, QD was viewed as a first-stage screener, while DA was conceptualized as a second-stage screener.

The first notable finding was that in the face of four competing predictors, QD did not emerge as a significant predictor of either outcome, for either group of children. We included QD because it is (a) a widely used screener in RTI systems; (b) a documented predictor of future mathematics performance (Chard et al., 2005; Clarke & Shinn, 2004; Lembke & Foegen, 2005; Seethaler & Fuchs, 2010b; Seethaler et al., 2012); and (c) not redundant with the transparent content of the CA or WP outcome. Given that QD predicts CA outcomes and that WP skill partially depends on CAs, we hypothesized that QD would explain variance in both types of mathematics outcomes. Yet three of the other four predictor measures were uniquely predictive, even with competing variables in the models. Thus, TEMA, DA, and vocabulary demonstrated superior predictive value compared to QD, and sole reliance on one of the alternative predictors (depending on outcome and student group) would produce greater accuracy in identifying risk for mathematics difficulty than would occur for exclusive reliance on QD.

Brief static screeners, such as QD, can produce unacceptably high rates of false positives (e.g., Seethaler & Fuchs, 2010), and reducing false positives would decrease the costs associated with multi-tier prevention systems by helping schools avoid costly intervention for students who do not require intervention. As mentioned, reducing false positives can be accomplished in an efficient manner with a gated screening system, in which an additional measure is administered to the subset of students who fail a brief, static, universal screen.

Present findings indicate that of the measures considered in this study, candidates for second-stage screening at first grade include TEMA, DA, and vocabulary. Results also suggest that the potential value of these measures depends on the mathematics outcome to be predicted. For the CA outcome, TEMA and DA each uniquely explained variance for LEP children, but follow-up tests indicated TEMA offered stronger predictive value than DA. On the other hand, for non-LEP children, TEMA was the only significant predictor. So despite the distinction in results for LEP versus non-LEP children, findings indicate that for predicting first-grade CA outcomes, TEMA may be the best choice for second-stage screening across LEP and non-LEP children. Deriving the same conclusion across LEP and non-LEP children about the optimal second-stage screener for predicting CA outcomes offers efficiency in the context of multi-tier support systems. Importantly for the purpose of the present study, it also provides support for screening first graders with LEP in the language of instruction (English).

At the same time, results do not provide the basis for recommending use of the same second-stage screener for formulating predictions across CA and WP outcomes, which is in line with previous research showing different constellations of predictors as a function of mathematics outcomes (e.g., Fuchs et al., 2010a,b; Seethaler et al., 2011). In forecasting WP outcomes for LEP children, TEMA was not significant. Instead, vocabulary and DA each uniquely explained variance, and follow-up tests showed stronger predictive value for DA. On the other hand, for non-LEP children, DA and TEMA were each uniquely predictive, without significant differences in their predictive value. Therefore, with the goal of training testers on only one second-stage screener for WP difficulty, DA appears to represent the best choice for second-stage screening across LEP and non-LEP children. Deriving the same conclusion across LEP and non-LEP children about the optimal second-stage screener for predicting WP outcomes again offers efficiency in the context of multi-tier support systems, while providing support for screening first graders with LEP in the language of instruction.

In the case of forecasting WP outcomes, it is notable that results support the potential value of DA, which indexes how well a child learns, rather than what the child already knows. This conclusion suggests the importance of additional research on dynamic assessment tools for forecasting students' learning potential and designating risk for poor outcomes. At the same time, findings raise the question, Why might DA provide strong value for predicting WP outcomes, while the extended static assessment emerged as the best predictor of CA outcomes?

With respect to TEMA's superior prediction of CA outcomes, we note that although the TEMA includes some WP items, its focus on CAs is stronger. It not only indexes skills foundational to first-grade addition and subtraction (e.g., counting, partitioning sets) but also directly assesses CA ideas and procedures with a variety of tasks: nonverbal addition, demonstration of addition with manipulatives, 2-digit addition and subtraction, timed number combinations, as well as understanding of the commutativity principle (which permits solution of subtraction problems with addition knowledge). By contrast, although this study's DA requires simple sums to 9, execution of those CAs is contextualized within a more complex, unfamiliar task (solving for missing numbers in mathematical expressions). Moreover, in the second and third instructional scaffolding levels, children are encouraged to rely on number lines for deriving CA solutions (which were not available during testing on the study's CA outcome), even as the children's attention is focused on the conceptual basis for missing numbers in mathematical expressions rather than on executing CAs. So, it is not surprising that TEMA proved stronger than DA in predicting CA outcomes, across LEP and non-LEP groups.

On the other hand, DA's greater explanatory value in predicting WP outcomes than CA outcomes makes sense for two reasons. First, DA addresses a common misunderstanding (e.g., Knuth et al., 2006) that is associated with WP (not CA) difficulty (e.g., Powell & Fuchs, 2010): interpreting the equal sign as a one-directional operator (input on the left results in output on the right of the equal sign). So it is not surprising that responsiveness to instruction on the DA task plays a specific role in predicting WP, not CA outcomes. Second, WP skill involves two forms of symbolic representations (numerals and language) and reflects understanding of relationships between known and unknown quantities. For example, in WP solving, the statement "Mary had 6 more than John," translated to M = 6 + J, by understanding that the smaller quantity must increase to equal the larger quantity. Yet, a common error is M + 6 = J. So WPs may rely more than CAs on the type of mental flexibility and manipulation of symbolic associations that is required in the DA's balancing equation tasks (Kieran, 1992; Sfard & Linchevski, 1994). We also note that balancing-equation DAs (similar to and different from the one used in the present study) have previously been shown to predict WP outcomes at third grade (Fuchs, Fuchs, Compton, et al., 2008) and to have stronger predictive value for WP than CA outcomes at first grade (Seethaler et al., 2011). (But these studies did not include LEP students.)

Because CAs have represented a dominant focus in mathematics research generally and screening studies specifically, some may question why we included WPs as a predicted outcome. Other than the obvious, transparent differences between CAs and WPs, it was important to include WPs as a distinct predicted outcome for the following reasons. We did so because WPs are the best school-age predictor of employment and wages in adulthood (e.g., Every Child a Chance Trust, 2009). Also, WPs can be a persistent deficit even when CA skill is adequate (Swanson et al., 2008), and the cognitive processes involved in WPs differ from and are more numerous than those underlying CAs (e.g., Fuchs, Geary, et al., 2010a,b; Fuchs et al., 2006; Fuchs, Fuchs, Stuebing et al., 2008). Since WPs and CAs may represent distinct components of mathematics competence, screening exclusively for CA outcomes may miss students who require intervention. This is the case at first grade, when WP skill represents a salient form of complex mathematical reasoning and when individual differences in WP competence grow rapidly (Fuchs et al., 2012).

Before closing, we caution readers to take note of study limitations. First, we considered only five predictor constructs: brief static domain-specific competence, more extended, comprehensive domain-specific competence, domain-general language ability (i.e., vocabulary), domain-general reasoning ability, and domain-specific learning potential (DA), each operationalized with a specific measure. It is possible that including additional or different constructs or operationalizing constructs with other measures would alter the pattern of results. For example, centering the DA tasks on more difficult calculations rather than on unfamiliar equation formats may have altered the predictive role of DA. Thus, additional research on this topic is clearly warranted. A second limitation is that distribution of the DA predictor and the WP outcome variables were positively skewed. (Because results for the original and the transformed variables were similar, we included results of the original variables for ease of interpretation.) Finally, in line with standard administration of QD, we did not award credit for correct pointing without the appropriate number label. That is, all students were required to name aloud the larger number (i.e., of pairs of numbers) to receive credit for correct responses. Because of this, we wonder if the scores of LEPs might have differed had we allowed for pointing in lieu of naming; it is possible that the role of QD as a predictor may have differed, as well. Future research on QD should consider and evaluate this possibility.

These issues notwithstanding, we offer the following conclusions. The potential value of different predictors of first-grade mathematics achievement depends on the outcome to be predicted and to a lesser extent on LEP status. In predicting CAs, TEMA and DA each uniquely explained variance for LEP children, with stronger predictive value for TEMA; for non-LEP children, TEMA was the only significant predictor. By contrast, in predicting WPs, only DA and vocabulary each uniquely explained variance for LEP children, with stronger predictive value for DA; for non-LEP children, DA and TEMA were uniquely predictive, without difference in their predictive value. Neither QD nor reasoning was significant in predicting either outcome. Results therefore provide the basis for considering a gated screening process, in which an extended mathematics assessment (such as TEMA) is used as a second-stage screener to predict CAs and DA is used as a second-stage screener to predict WPs. These conclusions warrant additional research using similar and different sets of predictor variables and outcome measures, conducted with other groups of LEP and non-LEP children.

References

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14,* 219-234. doi:10.1207/S15324818AME1403_2

Aiken, L.S., & West, S.G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Aud, S., Fox, M.A., & KewalRamani, A. (2010). *Status and trends in the education of racial and ethnic groups* (NCES 2010-015). Washington: U.S., Department of Education, National Center for Education Statistics. http://nces.ed.gove/pubs2010/2010015.pdf

Berk, L.E. (2009). *Child development* (8th ed.). Upper Saddle River, NJ: Pearson.

Bransford, J.C. , Delclos, V.R., Vye, N.J., Burns, M.S. & Hasselbring, T.S. (1987). State of the art and future directions. In C.S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potentia*l (pp. 479-496). New York: Guilford.

Budoff, M. (1967). Learning potential among institutionalized young adult retardates. *American Journal of Mental Deficiency*, 72, 404-411.

Campione, J.C. (1989). Assisted testing: A taxonomy of approaches and an outline of strengths and weaknesses. *Journal of Learning Disabilities, 22,* 151-165.

Campione, J.C., & Brown, A.L. (1987). Linking dynamic assessment with school achievement. In C. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential.* New York: Guilford.

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.

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. doi:10.1037/a0018488

Council of Chief State School Officers (2009). *English language development assessment.* Washington, DC: Author.

Ferrara, R.A., Brown, A.L., & Campione, J.C. (1986). Children’s learning and transfer of inductive reasoning rules: Studies of proximal development. *Child Development, 57,* 1087-1099.

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. doi:10.1037/a0027475

Fuchs, D., Fuchs, L.S., & Compton, D.L. (2012). Smart RTI: A next-generation approach to multi-level prevention. *Exceptional Children, 78,* 263-279.

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. doi:10.1037/a0012657

Fuchs, L.S., Fuchs, D., Compton, D.L., Powell, S.R., Seethaler, P.M., Capizzi, A.M., … Fletcher, J.M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98,* 29-43. doi:10.1037/0022-0663.98.1.29

Fuchs, L.S., Fuchs, D., Stuebing, K., Fletcher, J.M., Hamlett, C.L., & Lambert, W. (2008). Problem solving and computational skill: Are they shared or distinct aspects of mathematical cognition? *Journal of Educational Psychology, 100,* 30-47.

Fuchs, L.S., Geary, D.C., Compton, D.L., Fuchs, D., Hamlett, C.L., & Bryant, J. (2010a). The contributions of numerosity and domain-general abilities to school readiness. *Child Development, 81,* 1520-1533. doi:10.1111/j.1467-8624.2010.01489.x

Fuchs, L.S., Geary, D.C., Compton, D.L., Fuchs, D., Hamlett, C.L., Seethaler, P.M., … Schatschneider, C. (2010b). Do different types of school mathematics depend on different constellations of numerical versus general cognitive abilities? *Developmental Psychology, 46*, 1731-1746. doi:10.1037/a0020662

Feuerstein, R. (1979). *The dynamic assessment of retarded performers. The Learning Potential Assessment Device, theory, instruments, and techniques*. Baltimore: University Park.

Garcia, E.E., & Scribner, K.P. (2009). Latino Pre-K-3 education: A critical foundation. In E.E. Garcia & K.P. Scribner (Eds.), *Handbook of U.S. Latino psychology: Developmental and community-based perspectives* (pp. 267-289). Thousand Oaks, CA: Sage.

Geary, D.C., Bailey, D.H., & Hoard, 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. doi:10.1177/0734282908330592

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.

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

Grigorenko, E.L., & Sternberg, R.I. (1998). Dynamic testing. *Psychological Bulletin,* 124, 85-111.

Hofstetter, C.H. (2003). Contextual mathematics accommodation test effects for English-language learners. *Applied Measurement in Education, 16,* 159-188.

Holmbeck, G.N. (1997). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology, 65*, 599-610.

Jenkins, J.R., Hudson, R.F., & Johnson, E.S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review, 36,* 582-600.

Johnson, E.S., Jenkins, J.R., Petscher, Y., & Catts, H.W. (2009). How can we improve the accuracy of screening? *Learning Disabilities Research & Practice, 24*, 174-185.

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.

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.

Kieran, C. (1992). The learning and teaching of school algebra. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 390-419). Reston, VA: National Council of Teachers of Mathematics.

Knuth, E.J., Stephens, A.C., McNeil, N.M., & Alibali, M.W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education, 37*, 297-312.

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.

Lembke, E.S., & Foegen, A. (2009). Identifying early numeracy indicators for kindergarten and grade 1 students. *Learning Disabilities Research & Practice, 24,* 12-20.

Mancilla-Martinez, J., & Kieffer, M.J. (2010). Language minority learners' home language use is dynamic. *Educational Researcher, 39,* 545-546. doi:10.3102/0013189X10383168

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. doi:10.1207/s15327647jcd0602_6

Murray, B. A., Smith, K. A., & Murray, G. G. (2000). The test of phoneme identities: Predicting alphabetic insight in pre-alphabetic readers. *Journal of Literacy Research*, 32, 421-477.

Psychological Corporation. (1999). *Wechsler Abbreviated Scale of Intelligence.* San Antonio, TX: Harcourt Brace.

Research Institute on Progress Monitoring. (2009). *Early numeracy indicators.* Retrieved from http://www.progressmonitoring.org/pdf/TREarlymath6.pdf

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

Seethaler, P.M., & Fuchs, L.S. (2006). The cognitive correlates of computational estimation skill among third-grade students. *Learning Disabilities Research & Practice*, *21,* 233-243.

Seethaler, P.M., & Fuchs, L.S. (2010a). *Balancing Equations Dynamic Assessment.* Available from P.M. Seethaler, 228 Peabody College, Nashville, TN 37203.

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. (2011). Predicting first graders' development of calculation versus word-problem performance: The role of dynamic assessment. *Journal of Educational Psychology, 104,* 224-234. doi:10.1037/a0024968

Seethaler, P.M., Fuchs, L.S., Star, J.R., & Bryant, J. (2011). The cognitive predictors of computational skill with whole versus rational numbers: An exploratory study. *Learning and Individual Differences, 21,* 536-542. doi:10.1016/j.lindif.2011.05.002

Sfard, A., & Linchevski, L. (1994). The gains and the pitfalls of reification – the case of algebra. *Educational Studies in Mathematics, 26*, 191-228.

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

Spector, J.E. (1992). Predicting progress in beginning reading: Dynamic assessment of phonemic awareness. *Journal of Educational Psychology, 84,* 353-363.

Speece, D.L., Cooper, D.H., & Kibler, J.M. (1990). Dynamic assessment, individual differences, and academic achievement. *Learning and Individual Differences, 2*, 113-127.

Stansfield, C.W. (2011). Oral translation as a test accommodation for ELLs. *Language Testing, 28,* 401-416. doi:10.1177/0265532211404191

Sternberg, R.J. (1996). *Successful intelligence.* New York, NY: Simon & Schuster.

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.

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.* Nashville: Author.

Tzuriel, D., & Haywood, H.C. (1992). The development of interactive-dynamic approaches for assessment of learning potential. In H.C. Haywood and D. Tzuriel (Eds.), *Interactive assessment* (pp. 3-37). New York: Springer-Verlag.

U.S. Department of Commerce, Census Bureau. (2008). *Methodology for the United States Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2000 to July 1, 2008.*

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. doi:10.1016/j.jecp.2013.02.002

Vygotsky, L.S. (1962). *Thought and language.* Cambridge, MA: MIT Press.

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.

Wechsler, D. (1999). *Wechsler Intelligence Scale for Children.* San Antonio: Psychological Corporation.

Wilkinson, G.S. (1993). *Wide range achievement test 3.* Wilmington, DE: Wide Range.

Yates, J.R., & Ortiz, A.A. (2004). Classification issues in special education for English language learners. In A.M. Sorrels, H.J. Reith, & P.T. Sindelar (Eds.), *Critical issues in special education: Access, diversity, and accountability* (pp. 38-56). Boston: Pearson.

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.* Washington DC: Development Associates.

Zhu, J. (1999). *WASI manual.* San Antonio, TX: Psychological Corporation.

Table 1
Student Demographics by Group

	Group									
	Limited English Proficient (LEP) (n = 129)				Non-LEP (n = 163)					
Variable	*n*	%	*M*	*(SD)*	*n*	%	*M*	*(SD)*	χ^2	*F*
Sex									0.14	
Male	63	49			76	47				
Female	66	51			87	53				
Ethnicity/Race									201.47***	
Black	0	0			96	59				
White	0	0			37	23				
Hispanic	129	100			27	17				
Asian	0	0			1	1				
Biracial	0	0			2	1				
Subsidized Lunch									13.40***	
No	0	0			16	10				
Yes	129	100			147	90				
Special Education Services									7.87	
None	121	94			149	91				
Learning, Speech, Language, and/or Attention Deficit Disorder	8	6			14	9				
Years Retained, Previously										1.31
0	123	95			150	92				
1	6	5			13	8				
Age (in years)			6.6	(0.4)			6.6	(0.4)		3.33
Number Sets[a]			0.28	(1.1)			0.08	(1.1)		2.24

***p<.001. [a] Numbers Sets is *d* prime score for Number Sets Test (Geary et al., 2009).

Table 2

Means, Standard Deviations, and Correlations[a] Among Predictors and Outcome Measures for LEP[b] (n = 129) and Non-LEP (n = 163) Groups

	Raw score		Standard score[c]								
Measure	*M*	*(SD)*	*M*	*(SD)*	V	R	QD	T	DA	CA	WP
LEP											
Predictors											
Vocabulary (V)	10.91	(6.81)	31.76	(9.92)	--						
Reasoning (R)	6.97	(4.44)	45.95	(8.23)	.26	--					
Quantity Discrimination (QD)	21.80	(9.23)			.42	.27	--				
TEMA (T)	30.46	(8.40)	91.01	(12.11)	.48	.42	.63	--			
Dynamic Assessment (DA)	5.88	(4.53)			.39	.37	.52	.63	--		
Outcomes											
Calculations (CA)	17.48	(3.02)	96.91	(15.60)	.36	.31	.47	.66	.57	--	
Word Problems (WP)	3.96	(2.55)			.42	.22	.46	.51	.56	.52	--
Non-LEP											
Predictors											
Vocabulary (V)	15.99	(5.77)	38.84	(9.27)	--						
Reasoning (R)	7.20	(4.11)	45.90	(7.62)	.21	--					
Quantity Discrimination (QD)	25.33	(8.81)			.19	.29	--				
TEMA (T)	33.80	(7.21)	94.60	(10.48)	.28	.49	.66	--			
Dynamic Assessment (DA)	5.84	(4.51)			.27	.38	.46	.59	--		
Outcomes											
Calculations (CA)	17.95	(2.94)	96.82	(14.45)	.28	.31	.44	.69	.47	--	
Word Problems (WP)	4.60	(3.07)			.28	.42	.49	.62	.57	.58	--

Note. Vocabulary is Wechsler Abbreviated Scale of Intelligence (WASI) Vocabulary; Reasoning is WASI Matrix Reasoning; TEMA is Test of Early Mathematics Ability, 3rd ed.

[a]All correlations significant at $p < .01$ except for WP with R for LEP and QD with V for non-LEP, which were both significant at $p < .05$. [b]LEP is limited English proficient. [c]Standard scores for WASI Vocabulary and WASI Matrix Reasoning are *T* scores ($M = 50$; $SD = 10$); for Calculations (Wide-Range Achievement Test-Arithemetic), the mean is 100 ($SD = 15$).

Table 3

Regression Models Predicting Individual Differences in First-Grade Mathematics Development for LEP and Non-LEP Students

Outcome	*B*	*SE*	β	*t* (5, 123)	*p*	Outcome	*B*	*SE*	β	*t* (5, 157)	*p*
LEP						Non-LEP					
Calculations						Calculations					
Constant	10.91	0.79		13.87	<.001	Constant	8.45	0.92		9.23	<.001
Vocabulary	0.01	0.03	0.02	0.20	.85	Vocabulary	0.05	0.03	0.09	1.48	.14
Reasoning	0.00	0.05	-0.00	-0.01	.99	Reasoning	-0.04	0.05	-0.06	-0.82	.41
QD	0.01	0.03	0.03	0.36	.72	QD	-0.01	0.03	-0.04	-0.52	.60
TEMA	0.18	0.04	0.49	4.88	<.001	TEMA	0.27	0.04	0.65	7.25	<.001
DA	0.16	0.06	0.24	2.74	.01	DA	0.07	0.05	0.11	1.48	.14
Word Problems						Word Problems					
Constant	0.17	0.71		0.24	.81	Constant	-3.24	0.97		-3.34	<.001
Vocabulary	0.07	0.03	0.18	2.16	.03	Vocabulary	0.04	0.03	0.07	1.20	.23
Reasoning	-0.03	0.05	-0.06	-0.74	.46	Reasoning	0.09	0.05	0.12	1.75	.08
QD	0.03	0.03	0.12	1.26	.21	QD	0.04	0.03	0.11	1.36	.18
TEMA	0.05	0.03	0.15	1.43	.15	TEMA	0.14	0.04	0.32	3.48	<.001
DA	0.20	0.05	0.35	3.76	<.001	DA	0.18	0.05	0.27	3.63	<.001

Note. LEP is limited English language proficiency; Vocabulary is Wechsler Abbreviated Scale of Intelligence (WASI) Vocabulary; Reasoning is WASI Matrix Reasoning; QD = Quantity Discrimination; TEMA = Test of Early Mathematics Ability, 3rd ed.; DA = Dynamic Assessment.