ABSTRACT

HERNANDEZ, JESSICA ANN. A Meta-Analysis of STEM Interventions for Multilingual Learners. (Under the direction of Dr. Kate Norwalk).

Due to opportunity gaps, multilingual learners are at an increased risk of academic failure, grade retention, dropping out of school, and not attending higher education institutions (Locke & Sparks, 2019; Kanno & Cromley, 2015; Sheng et al., 2011). Interventions that could ameliorate these effects must be identified. Most empirical research has focused on literacy and reading interventions with multilingual learners. However, other subject areas such as science and math are equally as important. This study utilized a meta-analytic methodology to synthesize research findings on the effectiveness of Science, Technology, Engineering, and Mathematics (STEM) interventions with multilingual learners. After coding for inclusion and exclusion criteria, six science and five mathematics group-design studies were included in the final analysis. Mean effect sizes were found to be moderate for both science and math interventions, indicating that multilingual learners can increase their STEM achievement when provided useful intervention strategies. Effect sizes for science interventions varied by intervention type; interventions that were adult-directed *and* peer-mediated were found to have significantly larger mean effect sizes. Additionally, science and math interventions that incorporated culturally and linguistically responsive instruction practices were found to have significantly larger mean effect sizes. Overall, these results suggest there needs to be more research on multi-modal, culturally and linguistically responsive STEM interventions with multilingual learners because they can increase opportunities for academic success.

A Meta-Analysis of STEM Interventions for Multilingual Learners

by
Jessica Hernandez

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Psychology

Raleigh, North Carolina
2024

APPROVED BY:

Kate Norwalk, PhD
Committee Chair

Scott Stage, PhD

Jennie Hyde, PhD

Alison Turner, PhD

BIOGRAPHY

Jessica Hernandez is a School Psychology doctoral candidate at North Carolina State University. Prior to graduate school, she studied abroad in Perú and worked as a Special Education Instructional Assistant and as a Specialized Recreation and Inclusion Services Camp Counselor. As a psychologist-in-training, she has completed advanced practica on a multidisciplinary assessment team with Wake County Preschool Programs and as a therapist with Hope Services, LLC. She is currently an intern with the NC State Psychoeducational Clinic and New School Montessori Center in Holly Springs, NC. She is working towards certification as a Parent-Child Interaction therapist through the Center for Child and Family Health and is accredited as a Level 4 Group Triple P (Positive Parenting Program) provider. She is passionate about culturally responsive services and family, school, and community collaboration.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The number of people in the United States from diverse linguistic backgrounds increases daily (Colby & Ortman, 2014). This growth is reflected in the student population of U.S. public schools as multilingual learners continue entering the education system. Multilingual learners are students who learn English while speaking one or more languages and have the asset of multilingualism (Yankelowitz, 2023). Between 2000 and 2017, the number of multilingual learners in public schools increased in all but seven states and the District of Columbia to comprise more than 10% of the total student population (U.S. Department of Education, 2021), and this percentage is expected to continue rising. Due to international migration, this population increase will eventually exceed the natural birth rate in the United States (Vespa et al., 2018). Multilingual learners are a heterogenous group from many different cultures and speaking more than 400 languages (Language data, 2022); and yet, despite this growing diversity, scant empirical research exists on the effectiveness of specific instructional practices for multilingual learners (Cole, 2014), and many educators report feeling unprepared to instruct multilingual learners (Ballantyne et al., 2008).

Multilingual learners consistently lag behind their English-dominant counterparts on standardized tests of achievement and other academic outcomes (National Clearinghouse, 2015; Polat et al., 2016), placing them at a heightened risk for academic failure, grade retention, dropping out of school, and not attending higher education institutions (Locke & Sparks, 2019; Kanno & Cromley, 2015; Sheng et al., 2011). These difficulties highlight the need to identify interventions that successfully improve educational practices for multilingual learners. To date, most of the empirical research on instructional practices with multilingual learners has focused on literacy and reading (Cho et al., 2021; Cole, 2014; Hall et al., 2017). However, other content areas such as math and science need to be addressed. Thus, the purpose of this study was to synthesize research findings on Science, Technology, Engineering, and Mathematics (STEM) interventions with multilingual learners, examine which interventions are most effective, and identify areas for future research.

Opportunity Gaps

Much of the research on the academic achievement of multilingual learners has focused on a perceived achievement gap; however, many researchers and scholars have questioned the usefulness of identifying a perceived achievement gap between multilingual learners and students who identify as White, middle class (Foster, 1999; Ladson-Billings, 2006; Howard, 2010; Irvine, 2010). Criticisms of the perceived achievement gap have largely focused on the framing of White students as the intellectually and academically superior standard by which multilingual learners are compared, despite known cultural and linguistic biases in English language standardized tests. Additional criticisms of the perceived achievement gap include conceptualizing multilingual learners from a deficit perspective and focusing on individual or groups of students rather than the systems of inequity that lead to the perceived achievement gap (Milner, 2012). In reality, issues of opportunity are multifaceted and more nuanced than an achievement gap can explain. Irvine (2010) defines the perceived achievement gap as the culmination of other opportunity gaps (i.e., gaps in teacher quality, teacher training, challenging curriculum, school funding, a digital divide, wealth and income, employment opportunity, affordable housing, healthcare, nutrition, school integration, and quality childcare), which erroneously leads to the belief that an achievement gap truly exists. These opportunity gaps demonstrate some of the many reasons multilingual learners tend to score lower on state and federal tests of reading and mathematics (National Clearinghouse, 2015; Polat et al., 2016), placing them at a higher risk of dropping out of school (Sheng et al., 2011) and lowering the likelihood of going on to institutions of higher education (Kanno & Cromley, 2015).

Multilingual learners may experience a gap in their learning opportunities due to a lack of teacher preparation and access to interventions designed specifically for them. For teaching to be effective with multilingual learners, teachers must understand how to make curriculum culturally and linguistically meaningful, and they must be familiar with evidence-based approaches devised for instructing multilingual learners (Xu & Drame, 2007). Unfortunately, many teachers do not receive training in effective instruction for multilingual learners. Teachers may not be required to take any specific coursework related to the learning needs of multilingual learners, state teacher exams do not assess for knowledge or skills related to the learning needs of multilingual learners, and several states have no requirement regarding multilingual learners as part of their initial certification for all teachers (Samson & Collins, 2012). Due to lack of comfort or expertise, teachers may not be able or willing to implement recommended interventions with multilingual learners (Conway et al., 2000) or to evaluate intervention effectiveness (e.g., if instruction is appropriate or if a student should move to the next tier-level of intervention; Xu & Drame, 2007). Finally, teachers may not be aware of or have access to evidence-based interventions designed for multilingual learners. For example, a Florida school district designated a tier 2 reading intervention for all students but did not purchase the supplement for multilingual learners (Cavendish et al., 2016).

Addressing opportunity gaps that exist in educational practice is necessary for improving the academic achievement of multilingual learners. This meta-analysis seeks to investigate effective interventions/instructional practices in STEM for multilingual learners to better meet their needs.

The Importance of STEM Interventions for Multilingual learners

To date, the majority of research examining the effectiveness of academic interventions for multilingual learners has focused on literacy outcomes. For example, recent meta-analyses investigated the effectiveness of reading and literacy interventions with multilingual learners (e.g., Cho et al., 2021; Cole, 2014; Hall et al., 2017; Suggate, 2016), because reading is a significant predictor of overall school success (Guo et al., 2015). While this is important, other academic areas clearly need to be addressed. Results from the National Assessment of Educational Progress (NAEP) show there is an opportunity gap between multilingual learners and English-dominant peers not only in reading, but in mathematics (Polat et al., 2016) and science (Llosa et al., 2016) as well. Multilingual learners are underrepresented in higher education generally as well as in the STEM workforce. For example, Hispanic workers make up 18% of the U.S. workforce but represent only 14% of STEM workers (Okrent & Burke, 2021). Black and Hispanic STEM employees have pointed to a lack of quality education, a lack of encouragement to pursue STEM, and a lack of role models in the field as major reasons for underrepresentation (Funk & Parker, 2018). Other studies have identified ways to increase recruitment and retention of underrepresented groups in STEM, including exposure to culturally relevant STEM education/interventions and role models in STEM, especially at early ages (Avendano et al., 2018; Lynch et al., 2018). Further, exposure to role models in STEM has been shown to improve students' self-concept, sense of belonging, self-efficacy, and attitudes toward areas of underrepresentation, which may in turn improve their academic performance and career path (Hernandez et al., 2017). Increased interest and participation in STEM careers can lead to higher-paying occupations and constriction of the earnings gap (Lynch et al., 2018).

Like reading, it is imperative for foundational math knowledge and skills to be developed in early grades, because the content continues building on that foundation. Foundational math knowledge such as number relationships, basic operations, counting, and number sequencing has been shown to have predictive validity on academic achievement (Geary et al., 2013; Nguyen et al., 2016). Those who are not experts in the pedagogy of multilingual learners may view math as a universal language; however, students must use academic language in English to acquire mathematical computation and reasoning skills, such as organizing and communicating information, as well as reading and solving math problems, and this must happen before symbolic comprehension develops (Freeman, 2012). Many math terms and symbols are unfamiliar to learners (e.g., quotient, > greater than), and math terms may also be homonyms (e.g., scale); math terms are much more precise and nuanced than how words are used interpersonally (Freeman & Crawford, 2008). Additionally, math is taught in phrases (e.g., simplify the equation), a cause-and-effect text structure, and metaphors (e.g., the greater than sign is like a hungry alligator; Freeman, 2012). Therefore, obstacles arise for multilingual learners in math pedagogy due to linguistic and cultural differences.

Early research on science pedagogy for multilingual learners emphasized hands-on activities. The idea behind this approach was that effective instruction made science concrete and experiential (Llosa et al., 2016). Since then, researchers have argued for the need to go beyond experiential learning strategies. Multilingual learners and their peers must *use* language while being engaged in and learning about science and engineering (Lee et al., 2013). Developing science interventions and training teachers to meet multilingual learners' academic needs in science is a major concern in U.S. education; at the federal and state level, science is being integrated into more systems of accountability (Llosa et al., 2016). The Every Student Succeeds Act (ESSA) requires the administration of at least three science assessments, once in grade 3 through 8, once in grade 6 through 9, and once in grade 10 through 12. At least 14 states are using science tests to rate school quality or student success, and at least five states to measure schools academically (Klein, 2018).

Moderating Variables

A number of moderating variables may impact the effectiveness of STEM interventions for multilingual learners. What is known about moderating variables is largely based on meta-analyses of reading interventions. Previous meta-analyses on reading interventions with multilingual learners have examined grade-level, group size, intervention type (e.g., teacher-directed, peer-mediated, computer-based), intervention administrator (e.g., classroom teacher, trained interventionist, experimenter, tutor), and SES as moderating variables. However, there has not been much agreement among researchers regarding how and to what extent these factors may moderate the effectiveness of interventions for multilingual learners. Each of these moderating variables is discussed in greater detail below.

Student Variables

Previous research on academic interventions with English-dominant students has shown that intervention effectiveness can vary by grade-level. Suggate (2016) found that reading interventions with English-dominant students in grades 3 and 4 were significantly more effective than in grades 1 and 2. A possible explanation is that reading instruction is more effective when children have developed foundational reading skills to draw upon (Suggate, 2016). Other studies have shown larger intervention effects for upper elementary multilingual learners compared to those in middle and high school (Cho et al., 2021). Additionally, Cole (2014) found that multilingual middle school students made significantly smaller gains than elementary and high school multilingual learners in peer-mediated literacy interventions. Due to developmental differences, socialization may be particularly challenging for multilingual middle school students; middle school students are also asked to read denser texts (Cole, 2014). In sum, this research suggests that intervention effectiveness may vary depending on when the intervention is received by multilingual learners, and that receiving intervention in the upper elementary grades may be particularly effective. Therefore, it is hypothesized that interventions delivered in upper elementary grades will be most effective.

Other student-related variables that may influence intervention effectiveness are socioeconomic status and language. It is well-known that low socioeconomic status is related to poorer academic outcomes; some of the complex contributing factors include family and schools' limited access to funds and resources for supporting academic development (Donovan & Cross, 2002; Kieffer, 2010). In a meta-analysis on reading instruction with multilingual learners conducted by Hall and colleagues (2017), a substantial portion of participants came from low-income backgrounds. Therefore, results could only be generalized to this population. Not much is known about the needs of multilingual learners from other socioeconomic backgrounds. Multilingual learners from other socioeconomic backgrounds may attend schools with better funding, more qualified teachers, and adequate support for academic achievement, and thus may experience more successful interventions. Therefore, it is hypothesized that interventions delivered to participants coming from upper-class backgrounds will be most effective.

Given the diversity of languages spoken by multilingual learners, more research is also needed to determine the degree to which language may moderate the effectiveness of STEM interventions. Interventions delivered in English may be more effective with multilingual learners who speak languages similar to English (e.g., German, Dutch) versus students who speak languages with more linguistic distance (e.g., Mandarin). Further, multilingual learners' English language development progresses at different rates. Most states classify multilingual learners into one of five levels of English proficiency (Short et al., 2018). Interventions delivered in English may also be more effective for multilingual learners with higher levels of English proficiency. Therefore, it is hypothesized that interventions delivered to participants who speak languages similar to English and with higher levels of English proficiency will be most effective.

Culturally and Linguistically Responsive Instruction, Intervention Type, and Administrator

Researchers support the use of culturally responsive pedagogy to increase academic achievement among multilingual learners (Hernandez et al., 2013). Driver and Powell (2017) list the following as key elements of culturally and linguistically responsive instruction: (1) stating measurable learning goals, (2) peer discourse, (3) native language use, (4) graphic organizers and manipulatives, (5) using students' ideas and experiences, and (6) using examples relevant to students' lives, culture, and heritage. Culturally and linguistically responsive instructional strategies make information more accessible for multilingual learners. Therefore, it is hypothesized that interventions that incorporate these strategies will be more effective than those that do not.

The modalities in which interventions are delivered may also impact the effectiveness of interventions for multilingual learners. In a meta-analysis of literacy interventions, Cole (2014) found that peer-mediated literacy interventions with multilingual learners were more effective than teacher-centered instruction by nearly half a standard deviation. Peer-mediation allows for discussion and collaboration. Multilingual learners may learn more from their peers because they are comfortable asking them for help and are motivated to attend to what they say to develop social relationships (Bicais & Correia, 2009). Also, given that peer discourse is a component of culturally and linguistically responsive instruction, interventions that incorporate peer-mediation may be especially effective for multilingual learners. Therefore, it is hypothesized that peer-mediated interventions will be most effective.

An intervention administrator is the person or device that delivers the intervention, including classroom teachers, trained interventionists, experimenters, computers, and tutors. Suggate (2016) summarizes the literature on the well-documented finding that researcher-administered interventions typically have large effect sizes (Dignath & Buttner, 2008; Ehri et al., 2001), whereas computer-based interventions yield smaller effect sizes (Cheung & Slavin, 2012; Ehri et al., 2001). Live instruction may be best because the intervention administrator can ensure students are actively engaged with the material, whereas students may more easily tune out at a computer, rush through a program, or avoid difficult concepts (Aagaard, 2015). Therefore, it is hypothesized that interventions delivered by an individual(s) will be more effective than those delivered by a computer.

Group Size and Dosage

In U.S. schools, a student's risk of academic failure often determines the intensity of intervention they receive (Suggate, 2016). For example, one-on-one intervention is often reserved for students performing below 90% of their peers, whereas small groups are used for students between the 10th and 25th percentiles. However, limited research suggests the most effective reading interventions for multilingual learners are not individualized. For example, Cho and colleagues (2021) found the most effective reading interventions for multilingual learners are conducted in medium-sized groups of 6-15 students. Group interventions may be more effective because they allow for discussion and collaboration, which promotes language acquisition (Short et al., 2018). Therefore, it is hypothesized that interventions delivered in a group setting will be more effective than interventions delivered one-on-one.

Contradictory research has been reported regarding dosage of intervention instruction. Dosage refers to the total number of hours that treatment participants receive instruction (Hall et al., 2017). Neither meta-analysis by Cho and colleagues (2021) nor Hall and colleagues (2017) found that intervention dosage had a significant effect on reading ability of multilingual learners, despite previous findings that it was a significant moderator (Genesee et al., 2006; Wanzek et al., 2013). This may be due to ignoring the instructional dosage of the control group. More specifically, children in the control group may be receiving instruction of the same dosage as the treatment group, if the treatment group is receiving intervention in place of typical instruction. Suggate (2016) found that reading interventions with multilingual learners conducted in addition to typical instruction showed greater effects. There was no directional hypothesis for this moderating variable.

The Current Study

Multilingual learners must be able to access the general curriculum in U.S. public schools. It is legally mandated by the U.S. Department of Education Office for Civil Rights, and it is vital to their academic and overall well-being (U.S. Department of Justice & U.S. Department of Education, 2015). Previous meta-analyses on academic interventions with multilingual learners have focused primarily on literacy and reading; thus, the effectiveness of interventions in other content areas such as math and science is unclear. Understanding this would be a critical contribution to the literature on opportunity gaps between multilingual learners and their English-dominant peers. Interventions that build on cultural experiences and background knowledge and utilize effective instructional strategies for multilingual learners can increase opportunities for academic success. Therefore, the purpose of this study is to synthesize the findings of recent research on STEM interventions with multilingual learners, investigate which interventions are most effective, and identify areas for further research. The specific research questions are as follows: (1) What is the efficacy of STEM interventions for multilingual learners? (2) Does the effectiveness of STEM interventions vary depending on grade-level, socioeconomic status (SES), language, intervention type (e.g., teacher-directed, peer-mediated, computer-based), intervention administrator, group size, and/or dosage?

CHAPTER 2

METHODOLOGY

Search Strategy for Identifying Relevant Studies

Relevant articles for this meta-analysis were identified by searching the educational and psychological databases PsycINFO, ERIC, and Elsevier for studies published since January 2010. Many different terms are used to describe students who are multilingual. This paper chose to use multilingual learner because it is asset-based and person-centered (Yankelowitz, 2023). However, as language has only recently shifted to more asset-based terminology, this meta-analysis searched for previous terms commonly used to describe multilingual learners in the literature. Searches were conducted using the following descriptors: “English language learner,” “English as second language,” “Science,” “Math,” “STEM,” “Interventions,” and “Instruction” separately and in combination. Additionally, references from one previous meta-analysis of math interventions for multilingual learners (Arizmendi et al., 2021) were further examined. These search strategies yielded 116 published and unpublished articles from 2010 to 2022. Next, potentially relevant studies were identified by reading titles and abstracts. If an abstract seemed relevant, the full study was retrieved. Interlibrary Loan requests were made for studies that were not immediately available. The list of potential studies was further limited to empirical academic intervention studies, excluding curriculum and professional development studies. This refined the pool of potential articles to 45 with duplicates across databases removed. Finally, after retrieving the full text, studies were examined more closely and coded for inclusion and exclusion criteria.

Criteria for Inclusion of Studies

To be included in the meta-analysis, studies had to meet the following inclusion criteria: (a) involved multilingual learners in grades K-12; (b) included a school-based STEM intervention group; (c) used an experimental or quasi-experimental research design; and (d) provided the sample size, pre-test and post-test mean scores, and standard deviations for both treatment and control groups. Studies without control groups and single-group design studies were not included in the meta-analysis but are described in the results qualitatively. Additionally, studies had to be published in English; however, they could be published in any country. Studies' sample participants could be of all nationalities and linguistic backgrounds, if the target language was English, so comparisons with North American multilingual learners could be made. Unpublished manuscripts were included for review to control for publication bias (e.g., dissertations). However, no unpublished study met inclusion criteria.

Criteria for Exclusion of Studies

Studies were excluded from analyses for three reasons. First, studies of post-secondary students were excluded, as the focus of this study was on school-based interventions in grades K-12. Second, studies of students in special education and bilingual immersion programs were excluded so direct comparisons between multilingual learners in K-12 general education settings could be made. Finally, studies were excluded if results were not disaggregated for multilingual learners or if they did not provide enough information to calculate an effect size.

Data Analysis

Studies were coded for inclusion and exclusion criteria by the lead researcher and an academic advisor. The lead researcher coded all studies and created a code book, which was discussed with the academic advisor (see Appendix A). The code book details how to make decisions based on inclusion and exclusion criteria and potential moderating variables. The academic advisor independently coded studies using the code book. There were two rounds of practice with examples. Independent coding was compared and discussed to ensure consensus between coders. Minor disagreements were resolved through further discussion and subsequent agreement (Hertlein, 2014).

After all eligible studies were coded, effect sizes were calculated as Cohen's *d* using the "effect size for mean differences of groups with unequal sample size within a pre-post-control design" calculator on Psychometrica (https://www.psychometrica.de/effect_size.html). Additionally, Marfo and Okeyere (2019) provides an equation for calculating the variance of Cohen's *d*. The variance of Cohen's *d* was calculated for each study as:

$$V_d = ((n_1 + n_2)/(n_1 n_2)) + ((d^2)/(2(n_1+n_2)))$$

where n_1 and n_2 are sample sizes and *d* is Cohen's *d.* Main effect analyses were computed using a standard mean difference effect size. Heterogeneity was assessed using the *Q* statistic. Moderating variables included grade-level, intervention type (e.g., teacher-directed, peer-mediated, computer-based), intervention administrator, group size, dosage, and number of culturally responsive practices. Moderating variables were analyzed using meta-regression analysis. Each moderator was tested in a separate regression model due to the limited number of included studies. Tau-U was given or calculated using the Tau-U calculator available from singlecaseresearch.org for single-case design studies.

CHAPTER 3

RESULTS

What is the Efficacy of STEM Interventions for Multilingual Learners?

Science Achievement

Six group studies involving 1,533 multilingual learners engaged in a science intervention were included in the final analysis. These studies targeted broad science achievement and STEM knowledge. Table 1.1 provides a description of the included science studies. These studies were published between 2010-2021. The type of intervention was split between an adult-directed and peer-mediated intervention (Aguirre-Muñoz et al., 2021; August et al., 2010), adult-directed only (Bravo & Cervetti, 2014; Cervetti et al., 2015), and computer-based (Yu et al., 2021; Zheng et al., 2014). The classroom teacher delivered the intervention in every study. Group size was split between whole-group instruction (Aguirre-Muñoz et al., 2021; August et al., 2010; Cervetti et al., 2015) and individual instruction (Bravo & Cervetti, 2014; Yu et al., 2021; Zheng et al., 2014). Only two studies provided enough information to calculate dosage, or the number of hours participants received the intervention (August et al., 2010; Bravo & Cervetti, 2014). The remaining studies either did not make any mention of treatment duration or were too vague to estimate dosage. All studies were found to use at least one culturally and linguistically responsive instruction practice. The most common practice was the use of graphic organizers and manipulatives, found in four studies (Aguirre-Muñoz et al., 2021; August et al., 2010; Bravo & Cervetti, 2014; Zheng et al., 2014). The least used practice was providing relevant instructional examples to participants' daily lives, pop culture, and cultural heritage, found in just one study (Aguirre-Muñoz et al., 2021). One study included Vietnamese- and Arabic-speaking multilingual learners (Yu et al., 2021), and the other studies included Spanish-speaking multilingual learners.

Only one study reported participants' level of English language proficiency, but there was not enough information to calculate a mean level of proficiency (Aguirre-Muñoz et al., 2021). Two studies used criterion- or norm-referenced science outcome measures (Yu et al., 2021; Zheng et al., 2014; see Table 1.2).

A random effects model generated a mean effect size estimate of 0.57 ($SE = .19$, $p < .01$) across all studies of science interventions. According to Cohen's criterion, this effect size is moderate. A homogeneity statistic Q was computed to determine whether studies shared a common effect size. Heterogeneity was detected for this mean effect size, Q ($df = 7$) = 65.97, $p < .001$, suggesting significant differences in treatment effects across studies.

Science Vocabulary Knowledge. Two studies examined science vocabulary knowledge in addition to general science achievement (Bravo & Cervetti, 2014; Cervetti et al., 2015). One study examined science vocabulary knowledge exclusively (Van Orman et al., 2021), and therefore was not included in the final analysis. This study involved 34 Spanish-speaking multilingual learners in seventh grade. The classroom teacher delivered the intervention via whole-group instruction; the intervention was adult-directed and peer-mediated. Three culturally and linguistically responsive instruction practices were used: (1) facilitating oral discussions with students, (2) allowing use of native language, and (3) using graphic organizers and manipulatives. Cohen's d was the effect size reported, $d = 1.80$. This effect size was greater than Bravo & Cervetti (2014; $d = 0.82$) and Cervetti et al. (2015; $d = -0.14$). This is expected as the science intervention Van Orman and colleagues (2021) implemented targeted vocabulary development specifically, rather than measuring it as a secondary outcome.

No Control Group. Two studies were not included in the final analysis because they examined science vocabulary knowledge and general science achievement without the use of a control group (Ardasheva & Tretter, 2017; Jackson et al., 2020). Ardasheva and Tretter's (2017) study involved 61 multilingual learners in a science vocabulary intervention in ninth and tenth grade. Multilingual learners in this study were described as having English proficiency at the "entering" or "beginning" level. The classroom teacher delivered the intervention via whole-group and small-group instruction; the intervention was adult-directed and peer-mediated. Dosage was given and participants spent a range of 2.5-7.5 hours receiving the science intervention. Three culturally and linguistically responsive instruction practices were used: (1) facilitating oral discussions with students, (2) allowing use of native language, and (3) using graphic organizers and manipulatives. Cohen's *d* was the effect size reported, $d = 0.59$.

Jackson and colleagues' (2020) study involved 128 Spanish-speaking multilingual learners across two schools in a science intervention. Their classroom teachers implemented the intervention via whole group instruction; the intervention was adult-directed. Only one culturally and linguistically responsive instruction practice was used – facilitating oral discussions with students. Cohen's *d* was the effect size reported. Cohen's *d* in the first year of implementation for school one was $d = 0.8$. Cohen's *d* in the first year of implementation for school two was $d = 0.76$.

Mathematics

Five group studies involving 9,484 multilingual learners engaged in a mathematics intervention were included in the final analysis. These studies targeted broad mathematics achievement, early numeracy, and word problem solving. Table 2.1 provides a description of the mathematics group-design studies. These studies were published between 2012-2019. The type of intervention was split between adult-directed (Doabler et al., 2019; Orosco & Abdulrahim, 2018) and computer-based (Crawford, 2013; Freeman, 2012; Rutherford et al., 2014). Three studies had a classroom teacher deliver the intervention (Crawford, 2013; Freeman, 2012; Rutherford et al., 2014), one study used an interventionist (Doabler et al., 2019), and one study used researcher-trained individuals (Orosco & Abdulrahim, 2018). Group size was split between small group (2-5 students; Doabler et al., 2019; Orosco & Abdulrahim, 2018) and individual instruction (Crawford, 2013; Freeman, 2012; Rutherford et al., 2014). Dosage was given or calculated for each study and participants spent a range of 15-60 hours receiving the mathematics intervention. Only two studies were found to use culturally and linguistically responsive instructional practices (Crawford, 2013; Freeman, 2012). Both studies allowed the use of native language and employed the use of graphic organizers and manipulatives. All studies included Spanish-speaking multilingual learners. Two studies additionally included Arabic-, Portuguese-, and Vietnamese-speaking individuals (Doabler et al., 2019; Rutherford et al., 2014). Three studies reported participants' level of English proficiency, with two providing enough information to calculate a mean proficiency level (higher numbers signify higher levels of proficiency; Doabler et al., 2019; Orosco & Abdulrahim, 2018). Three studies used norm-referenced mathematics outcome measures (Doabler et al., 2019; Orosco & Abdulrahim, 2018; Rutherford et al., 2014; see Table 2.2).

A random effects model generated a mean effect size estimate of 0.48 (SE = .16, $p < .01$). According to Cohen's criterion, this effect size is moderate. A homogeneity statistic Q was computed to determine whether studies shared a common effect size. Heterogeneity was detected for this mean effect size, Q (df = 7) = 87.22, $p < .001$, suggesting significant differences in treatment effects across studies.

Single-Subject Design Studies. Six single-subject design studies examined word problem-solving among multilingual learners. Table 3.1 provides a description of the single-subject design studies. These studies were published between 2013-2020. All interventions were adult-directed except one, which was computer-based (Xin et al., 2020). Intervention administrator was split between experimenters (Kim et al., 2015; Luevano & Collins, 2020; Orosco et al., 2013), researcher-trained individuals (Kong & Swanson, 2019; Luevano & Collins, 2020), a classroom teacher (Orosco, 2014), and interventionists (Orosco et al., 2013; Xin et al., 2020). Group size was split between individual instruction (Kim et al., 2015; Orosco et al., 2013; Xin et al., 2020), small group instruction (Kong & Swanson, 2019), and both (Luevano & Collins, 2020; Orosco, 2014). Dosage was given or calculated for five of the studies and participants spent a range of 4.92-11.67 hours receiving mathematics intervention (Kong & Swanson, 2019; Luevano & Collins, 2020; Orosco, 2014; Orosco et al., 2013; Xin et al., 2020). Three studies were found to use one culturally and linguistically responsive instruction practice (Kim et al., 2015; Orosco, 2014; Xin et al., 2020) and one study used two practices (Luevano & Collins, 2020). Four studies included Spanish-speaking multilingual learners, one study included Spanish- and Indonesian-speaking students (Kong & Swanson, 2019), and one study included Chinese- and Korean-speaking students (Kim et al., 2015). Three studies reported participants' level of English proficiency, with two providing enough information to calculate a mean proficiency level (Kong & Swanson, 2019; Luevano & Collins, 2020). One study used a norm-referenced mathematics outcome measure (Kong & Swanson, 2019; see Table 3.2).

For each study, Tau-U was reported or calculated to be over .45, which indicates strong positive associations. An average Tau-U was calculated for two studies that reported more than one outcome (Kim et al., 2015; Kong & Swanson, 2019), and then an average Tau-U was calculated across all studies. The average Tau-U estimate across studies was .76, suggesting the math interventions were effective.

Does the Effectiveness of STEM Interventions Vary Depending on Moderating Variables?

Science

Among studies included in the final analysis, separate meta-regression models were conducted for each moderating variable due to the limited number of included studies. The magnitude of effect size for science studies varied as a function of intervention type and the number of culturally and linguistically responsive practices used. Studies that used adult-directed *and* peer-mediated interventions showed a significantly larger mean effect size (.81, $p < .05$) than adult-only or computer-only interventions. The Q statistic was significant for intervention type, Q (df = 5) = 60.3, $p < .001$. The number of culturally and linguistically responsive practices used naturally fell into clear groupings and was therefore dichotomized into two groups. Studies that included four or more culturally and linguistically responsive practices showed a significantly larger mean effect size (.81, $p < .05$) than studies that used fewer than four. The Q statistic was significant for culturally and linguistically responsive practices, Q (df = 5) = 63.5, $p < .001$. This result should be interpreted with caution given the same two studies yielded both of these findings. Meta-regression models were not significant for grade-level or group size. Dosage was not given or could not be calculated for enough studies to be examined as a moderator. Additionally, there was not enough variability between studies to examine SES, intervention administrator, or language as moderators.

Mathematics

Similar to science interventions, the magnitude of effect size for math interventions varied between studies as a function of the number of culturally and linguistically responsive practices used. The number of culturally and linguistically responsive practices used naturally fell into clear groupings and was therefore dichotomized into two groups. Studies that included two or more culturally and linguistically responsive practices showed a significantly larger mean effect size (.71, $p = .05$) than studies that used none. The *Q* statistic was significant for culturally and linguistically responsive practices, Q (df = 3) = 24.97, $p < .001$. Meta-regression models were not statistically significant for grade-level, intervention administrator, intervention type, group size, or dosage. There was not enough variability between studies to examine SES or language as moderators.

CHAPTER 4

DISCUSSION

Multilingual learners are underrepresented in the STEM workforce, and culturally relevant STEM interventions have been identified as one way of increasing the recruitment and retention of underrepresented groups in STEM (Avendano et al., 2018; Lynch et al., 2018). Whereas previous meta-analyses of academic interventions with multilingual learners have primarily focused on reading and literacy, this study aimed to synthesize recent research on STEM interventions with multilingual learners, identify effective interventions for these students, and identify areas for further research. Overall, science and math interventions yielded moderate effect sizes. The magnitude of effect sizes varied between studies as a function of intervention type and the number of culturally and linguistically responsive practices used. These findings suggest that multilingual learners' opportunities for academic success can be increased with multi-modal interventions that utilize culturally and linguistically responsive instruction practices.

Effectiveness of STEM Interventions with Multilingual Learners

The primary goal of this study was to estimate the average effect size of STEM interventions with multilingual learners. Mean effect sizes were found to be moderate for both science and math interventions from the few studies included in this meta-analysis. Effect sizes of math single-subject design studies also showed strong positive associations. This indicates that multilingual learners can increase their STEM achievement when provided useful intervention strategies. This finding is also commensurate with previous meta-analyses of literacy interventions with multilingual learners (Cole, 2014; Cho et al., 2021), which demonstrated moderate mean effect sizes. Taken together, these findings support the premise that

STEM interventions for multilingual learners are needed, and schools should be investing in STEM interventions that utilize effective instructional strategies for multilingual learners because they can increase opportunities for academic success.

Despite these promising findings, the present study highlights that limited research on STEM interventions with multilingual learners exists, especially compared to studies of literacy interventions. For example, entering this study's search criteria into Google Scholar yielded 47,600 hits, and there were 189,000 hits when the search criteria were changed to reading and literacy. Research with multilingual learners has likely focused on literacy outcomes because reading is a significant predictor of overall school success (Guo et al., 2015). More empirical studies of STEM interventions with multilingual learners are needed given that gaps in opportunities to learn also exist in math and science. Additionally, culturally relevant STEM interventions have been identified as ways to increase representation in the STEM workforce and more equal distribution of income (Avendano et al., 2018; Lynch et al., 2018).

This meta-analysis also highlights the lack of empirical research into STEM interventions with multilingual learners, which led to many studies being excluded. After coding for inclusion and exclusion criteria, only six studies of science interventions and five of mathematics interventions were included in the final analysis. For example, studies that did not include a control group were excluded from the final analysis and studies that implemented an intervention but did not measure if that intervention was effective were excluded from this meta-analysis. More systematic research on STEM interventions with multilingual learners is needed, such as randomized control trials, to draw stronger conclusions about intervention effectiveness. Additionally, limited reporting of essential information led to the exclusion of several studies, including eight studies that did not disaggregate results for English proficient and multilingual learners. Likewise, 19 studies that did not report enough information to calculate an effect size were excluded. Results must be disaggregated for multilingual participants, and information to calculate effect sizes must be reported to conduct a meta-analysis on multilingual learners.

Moderation Analysis

A second purpose of this study was to determine if effect sizes for STEM interventions with multilingual learners varied depending on several moderating variables, including grade-level, SES, language, intervention type, intervention administrator, group size, and dosage. Results suggest that science interventions that are adult-directed *and* peer-mediated may be more effective for multilingual learners than adult-directed only and computer-based interventions. A multi-modal approach may be most effective due to multilingual learners being more comfortable asking peers for help, as well as being more motivated to work with peers to build their social network (Bicais & Correia, 2009). Multilingual learners may also rush, skip, or tune out when using a computer program (Aagaard, 2015). However, it is hard to tell if these studies were more effective due to intervention type or due to incorporating culturally and linguistically responsive instruction practices. Previous meta-analyses of literacy interventions with multilingual learners have found either adult-directed *or* peer-mediated interventions to be more effective than the other. It would be interesting to see if adult-directed *and* peer-mediated literacy interventions were most effective. Similarly, future math intervention studies should investigate adult-directed *and* peer-mediated interventions, as the five mathematics studies included in this meta-analysis were either adult-directed or computer-based.

Math *and* science interventions that used more culturally and linguistically responsive instruction practices were found to be most effective. This is not surprising because culturally and linguistically responsive pedagogy with multilingual learners is supported by research

(Hernandez et al., 2013). These practices make curriculum meaningful and therefore make information more accessible for multilingual learners (Xu & Drame, 2007).

Contrary to hypotheses, intervention efficacy did not vary based on grade-level or group size for math and science intervention studies. Additionally, intervention efficacy did not vary based on intervention type, intervention administrator, or dosage among math intervention studies. This is surprising because prior research in both literacy and STEM with multilingual learners has consistently found these variables to moderate intervention effectiveness (Arizmendi et al., 2021; Cho et al., 2021; Cole, 2014; Hall et al., 2017). It is likely that there was not enough power to detect statistically significant moderating effects for these variables due to the limited number of studies included in this meta-analysis.

Recommendations for Practice

A troubling finding is that most math group studies did not incorporate culturally responsive instructional practices at all. Of the math and science studies that did incorporate responsive instruction, the majority only used one or two strategies, which is insufficient for promoting meaningful participation for multilingual learners. Most science studies only employed the use of graphic organizers and manipulatives, and the least used practice was providing relevant instructional examples to participants' daily lives, pop culture, and cultural heritage. This is a missed opportunity to capitalize on students' assets that they bring into the classroom.

These findings are in direct contrast to *The 6 Principles for Exemplary Teaching of English Learners* published by TESOL International Association. The 6 Principles are evidence-based guidelines for teachers of multilingual learners in K-12 classrooms. The first three principles are relevant to the culturally and linguistically responsive instruction practices that were coded in this study. The first principle, at the core of good instruction, is “Know your learners.” This means teachers know their students’ families, languages, cultures, and assets, and draw on this information to prepare and deliver engaging and effective lessons (Short et al., 2018). For example, teachers can find out what background knowledge students already have about a topic when they instruct multilingual learners to speak with peers in their native language about the topic and report back in English, share what they know in their native language, draw what they know, or use a graphic organizer; teachers can build on this background knowledge with pictures, tangible objects, relating material to students’ personal experiences, and pre-teaching important vocabulary (Breiseth, 2021).

The second principle is “Create conditions for language learning.” This means teachers make decisions about the physical environment of the classroom, lesson materials, and social settings to ensure student comfort (Short et al., 2018). For example, teachers can pre-teach key phrases and new vocabulary, demonstrate that vocabulary has multiple meanings, explain the different meanings, and how to use vocabulary correctly in the context of STEM, encourage students to explain the material to another student in their native language, and demonstrate vocabulary through visuals, graphic representations, gestures, and tangible objects (Robertson, 2020).

The third principle is “Design high-quality lessons for language development.” This means teachers plan lessons that not only help students learn new strategies and skills, but that also promote language acquisition (Short et al., 2018). For example, teachers can display sentence frames, have students share how they came to their answer, have students answer each other’s questions, use journal prompts to have students process their learning, and challenge students to create their own problems (Robertson, 2020).

Ideally, interventions designed for multilingual learners are utilizing all 6 Principles in tandem. Unfortunately, results suggest this is not the case for now, but something to aspire to. Similar to the coding methodology of this meta-analysis, educators and school-based practitioners can evaluate interventions with multilingual learners to see if using students' own ideas and experiences, providing relevant instructional examples to participants' daily lives, pop culture, and cultural heritage, allowing the use of native language, using graphic organizers and manipulatives, explicitly stating measurable lesson objectives, and facilitating oral discussions with students are in place.

Limitations and Future Directions

This study adds to the limited research available on the effectiveness of STEM interventions for improving math and science achievement outcomes for multilingual learners. However, the findings must be interpreted in light of some limitations. Most notably, the reduced number of studies that met inclusion criteria limited the generalizability of the findings and made it difficult to explore independent and interactive effects of moderating variables. More empirical studies of STEM interventions with multilingual learners are needed in general so that more comprehensive meta-analyses can be conducted. In addition, this meta-analysis also highlights limitations in the methodology and reporting of studies themselves. Future meta-analyses may want to broaden their inclusion criteria to post-secondary students, students in special education, and/or bilingual immersion programs. Studies also frequently failed to report information on participants' English language proficiency and the amount of time students spent in intervention (i.e., dosage). Future studies should report English proficiency level because interventions delivered in English may be more effective for multilingual learners with higher levels of English proficiency. Finally, future studies should also report the total number of hours that treatment participants receive instruction to ensure they are receiving intervention in addition to typical instruction, and therefore getting treatment above and beyond the control group (Suggate, 2016).

This study was also limited by the participant characteristics in the included studies. Previous intervention research with multilingual learners, including the studies in this meta-analysis, have centered Spanish-speaking students. Given the diversity of languages spoken by multilingual learners, participants who speak languages other than Spanish should be recruited for future studies. Interventions delivered in English may be more effective with multilingual learners who speak languages similar to English (e.g., German, Dutch) versus students who speak languages with more linguistic distance (e.g., Mandarin). Likewise, previous intervention research with multilingual learners has focused on multilingual learners from low-income backgrounds. It is well-established that low socioeconomic status is related to poorer academic outcomes, but not much is known about the needs of multilingual learners from other socioeconomic backgrounds (Hall et al., 2017).

Conclusion

This study utilized a meta-analytic methodology to synthesize research findings on Science, Technology, Engineering, and Mathematics (STEM) interventions with multilingual learners. It is important to identify effective STEM interventions for multilingual learners to mitigate academic difficulties due to opportunity gaps. The results of this meta-analysis demonstrate research is largely lacking, and interventions with multilingual learners are failing to incorporate culturally and linguistically responsive instruction practices. Researchers must do better at implementing academic interventions with cultural responsiveness and reporting necessary information for determining intervention effectiveness with multilingual learners to help ensure the delivery of appropriate instruction for all.

REFERENCES

Aagaard, J. (2015). Drawn to distraction: A qualitative study of off-task use of educational technology. *Computers & Education*, *87*, 90-97. https://doi.org/10/1016/j.compedu.2015.03.010

Aguirre-Muñoz, Z., Pantoya, M., Pando, M., & Loria Garro, E. S. (2021). Engineering Integration in Elementary Science Classrooms: Effects of Disciplinary Language Scaffolds on English Learners' Content Learning and Engineering Identity. *Journal of Engineering Education*, *110*, 517–544. https://doi.org/10.1002/jee.20409

Ardasheva, Y., & Tretter, T. R. (2017). Developing science-specific, technical vocabulary of high school newcomer English learners. *International Journal of Bilingual Education and Bilingualism*, *20*, 252-271. https://doi.org/10.1080/13670050.2015.1042356

Arizmendi, G. D., Li, J., Van Horn, M. L., Petcu, S. D., & Swanson, H. L. (2021). Language-focused interventions on math performance for English learners: A selective meta-analysis of the literature. *Learning Disabilities Research & Practice*, *36*, 56-75. https://doi.org/10.1111/ldrp.12239

August, D., Branum-Martin, L., Cardenas-Hagan, E., & Francis, D. (2010). The Impact of an Instructional Intervention on the Science and Language Learning of Middle Grade English Language Learners. *Journal of Research on Educational Effectiveness*, *2*, 345-376. https://doi.org/10.1080/19345740903217623

Avendano, L., Renteria, J., Kwon, S., & Hamdan, K. (2018). Bringing equity to underserved communities through STEM education: implications for leadership development. *Journal of Educational Administration and History*, *51*, 66-82. https://doi.org/10.1080/00220620.2018.1532397

Ballantyne, K. G., Sanderman, A. R., & Levy, J. (2008). *Educating English language learners: Building teacher capacity*. National Clearinghouse for English Language Acquisition. Retrieved November 2, 2021, from https://files.eric.ed.gov/fulltext/ED535608.pdf

Bicais, J., & Correia, M. G. (2009). Peer-learning spaces: A staple in English language learners' tool kit for developing language and literacy. *Journal of Research in Childhood Education*, *22*, 363-375. https://doi.org/10.1080/02568540809594633

Bravo, M. A., & Cervetti, G. N. (2014). Attending to the language and literacy needs of English learners in science. *Equity & Excellence in Education*, *47*, 230-245. https://doi.org/10.1080/10665684.2014.900418

Breiseth, L. (2021). *Background knowledge and ELLs: What teachers need to know*. Colorín Colorado. https://www.colorincolorado.org/article/background-knowledge-and-ells-what-teachers-need-know

Cavendish, W., Harry, B., Menda, A. M., Espinosa, A., & Mahotiere, M. (2016). Implementing response to intervention: Challenges of diversity and system change in a high-stakes environment. *Teachers College Record*, *118*, 1-36. https://doi.org/10.1177/016146811611800505

Cervetti, G. N., Kulikowich, J. M., & Bravo, M. A. (2015). The effects of educative curriculum materials on teachers' use of instructional strategies for English language learners in science and on student learning. *Contemporary Educational Psychology*, *40*, 86-98. https://doi.org/10.1016/j.cedpsych.2014.10.005

Cheung, A. C., & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review*, *7*, 198-215. https://doi.org/10.1016/j.edurev.2012.05.002

Cho, Y., Kim, D., & Jeong, K. (2021). Evidence-based reading interventions for English language learners: A multilevel meta-analysis. *Heliyon*, *7*, 1-11. https://doi.org/10.1016/j.heliyon.2021.e07985

Colby, S. L., & Ortman, J. M. (2014). *Projections of the size and composition of the U.S. population: 2014 to 2060*. Current Population Reports (pp. 25-1143) U.S. Census Bureau.

Cole, M. W. (2014). Speaking to read: Meta-analysis of peer-mediated learning for English language learners. *Journal of Literacy Research*, *46*, 358-382. https://doi.org/10.1177/1086296X14552179

Conway, D. F., Christensen, J. E., Russell, J. F., & Brown, J. D. (2000). Principals' perceptions of pre-referral student assistance teams. *ERS Spectrum*, *18*, 14-19.

Crawford, L. (2013). Effects of an Online Mathematics Curriculum for English Language Learners. *Computers in the Schools*, *30*, 248-270. https://doi.org/10.1080/07380569.2013.805665

Dignath, C., & Buttner, G. (2008). Components of fostering self-regulated learning among students: A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, *3*, 231-264. https://doi.org/10.1007/s11409-008-9029-x

Donovan, M. S., & Cross, C. T. (2002). *Minority students in special and gifted education*. National Academies Press.

Doabler, C. T., Clarke, B., Kosty, D., Smolkowski, K., Kurtz-Nelson, E., Fien, H., & Baker, S. K. (2019). Building number sense among English learners: A multisite randomized controlled trial of a Tier 2 kindergarten mathematics intervention. *Early Childhood Research Quarterly*, *47*, 432-444. https://doi.org/10.1016/j.ecresq.2018.08.004

Driver, M. K., & Powell, S. R. (2017). Culturally and linguistically responsive schema intervention: Improving word problem solving for English language learners with mathematics difficulty. *Learning Disability Quarterly*, *40*, 41-53. https://www.jstor.org/stable/44280675

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly*, *36*, 250-287. https://doi.org/10.1598/RRQ.36.3.2

Foster, M. (1999). Race, class, and gender in education research: Surveying the political terrain. *Educational Policy*, *13*, 77-85. Retrieved May 10, 2024, from https://journals.sagepub.com/doi/pdf/10.1177/0895904899131007?casa_token=1k8cruN7cO0AAAAA:4wB1fL4pVH_ofLzAAbzzuAaZUr7kXB0ZWYpFwqd4Eet_vQUhLXjgCJCneWam8jfIej5wgepMhj9m4Q4

Freeman, B. (2012). Using digital technologies to redress inequities for English language learners in the English speaking mathematics classroom. *Computers & Education*, *59*, 50-62. https://doi.org/10.1016/j.compedu.2011.11.003

Freeman, B., & Crawford, L. (2008). Creating a middle school mathematics curriculum for English-language learners. *Remedial and Special Education*, *29*, 9-19. https://doi.org/10.1177/0741932507309717

Funk, C., & Parker, K. (2018). *Blacks in STEM jobs are especially concerned about diversity and discrimination in the workplace*. Pew Research Center. https://www.pewresearch.org/social-trends/2018/01/09/blacks-in-stem-jobs-are-especially-concerned-about-diversity-and-discrimination-in-the-workplace/

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2013). Adolescents' functional numeracy is predicted by their school entry number system knowledge. *PLoS ONE*, *8*, 1-8. https://doi.org/10.1371/journal.pone.0054651

Genesee, F., Lindholm-Leary, K., Saundrs, W. M., & Christian, D. (2006). *Education English language learners: A synthesis of research evidence*. Cambridge University Press.

Guo, Y., Sun, S., Breit-Smith, A., Morrison, F. J., & Connor, C. M. (2015). Behavioral engagement and reading achievement in elementary-school-age children: A longitudinal cross-lagged analysis. *Journal of Educational Psychology*, *107*, 332-347. https://doi.org/10.1037/a0037638

Hall, C., Roberts, G. J., Cho, E., McCulley, L. V., Carroll, M., & Vaughn, S. (2017). Reading instruction for English learners in the middle grades: A meta-analysis. *Educational Psychology Review*, *29*, 763-794. https://doi.org/10.1007/s10648-016-9372-4

Hernandez, C. C., Morales, A., & Shroyer, M. (2013). The development of a model of culturally responsive science and mathematics teaching. *Cultural Studies of Science Education*, *8*, 803-820. https://doi.org/10.1007/s11422-013-9544-1

Hernandez, D., Rana, S., Rao, A., & Usselman, M. (2017). Dismantling stereotypes about Latinos in STEM. *Hispanic Journal of Behavioral Sciences*, *39*, 436-451. https://doi.org/10.1177/0739986317731100

Hertlein, K. (2014). Advantages and disadvantages of technology in relationships: Findings from an open-ended survey. *The Qualitative Report*, 19, 1-11. Retrieved August 12, 2024, from https://nsuworks.nova.edu/tqr/vol19/iss11/2/

Howard, T. C. (2010). *Why race and culture matter in schools: Closing the achievement gap in America's classrooms*. New York, NY: Teachers College Press.

Irvine, J. J. (2010). Foreword. In H. R. Milner's (Ed.), *Culture, curriculum, and identity in*

Education (pp. xi-xvi). Palgrave Macmillan.

Jackson, J. K., Huerta, M., & Garza, T. (2020). A promising science and literacy instructional model with Hispanic fifth grade students. *The Journal of Educational Research*, *113*, 79-92. https://doi-org/10.1080/00220671.2020.1728734

Kanno, Y., & Cromley, J. G. (2015). English language learners' pathways to four-year colleges. *Teachers College Record: The Voice of Scholarship in Education*, *117*, 1-44. https://doi.org/10.1177/016146811511701202

Kieffer, M. J. (2010). Socioeconomic status, English proficiency, and late-emerging reading difficulties. *Educational Researcher*, *39*, 484-486. https://doi.org/10.3102/0013189X10378400

Kim, S. A., Wang, P., & Michaels, C. A. (2015). Using explicit CRA in- struction to teach fraction word problem solving to low-performing Asian English learners. *Reading & Writing Quarterly*, *31*, 253-278. https://doi.org/10.1080/10573569.2015.1030999

Klein, A. (2018). *How does science testing work under ESSA?* EducationWeek. https://www.edweek.org/policy-politics/how-does-science-testing-work-under-essa/2018/10

Kong, J. E., & Swanson, H. L. (2019). The effects of a paraphrasing intervention on word problem-solving accuracy of English learners at risk of mathematic disabilities. *Learning Disability Quarterly*, *42*, 92-104. https://doi.org/10.1177/0731948718806659

Ladson-Billings, G. (2006). From the achievement gap to the educational debt: Understanding achievement in U.S. schools. *Educational Researcher*, *35*, 3-12. https://doi.org/10.3102/0013189X035007003

Language data for the United States of America. (2022). *Translators without Borders*.

Retrieved October 20, 2022, from https://translatorswithoutborders.org/language-data-for-the-united-states-of-america

Lee, O., Quinn, H., & Valdés, G. (2013). Science and language for English language learners in relation to Next Generation Science Standards and with implications for Common Core State Standards for English language arts and mathematics. *Educational Researcher*, *42*, 411-438. https://doi.org/10.3102/0013189X13480524

Llosa, L., Lee, O., Jiang, F., Haas, A., O'Connor, C., Van Booven, C. D., Kieffer, M. J. (2016). Impact of a large-scale science intervention focused on English language learners. *American Educational Research Journal*, *53*, 395-424. https://doi.org/10.3102/0002831216637348

Locke, V. N., & Sparks, P. J. (2019). Who gets held back? An analysis of grade retention using stratified frailty models. *Population Research and Policy Review*, *38*, 695-731. https://doi.org/10.1007/s11113-019-09524-3

Luevano, C., & Collins, T. A. (2020). Culturally appropriate math problem-solving instruction with English language learners. *School Psychology Review*, *49*, 144-160. https://doi.org/10.1080/2372966X.2020.1717243

Lynch, S. J., Burton, E. P., Behrend, T., House, A., Ford, M., Spillane, N., & Means, B. (2018). Understanding inclusive STEM high schools as opportunity structures for underrepresented students: Critical components. *Journal of Research in Science Teaching*, *55*, 712-748. https://doi.org/10.1002/tea.21437

Marfo, P., & Okyere, G. A. (2019). The accuracy of effect-size estimates under normal and contaminated normals in meta-analysis. *Heliyon*, *5*, Article e01838. https://doi.org/10.1016/j.heliyon.2019.e01838

Milner, H. R. (2012). Beyond a test score: Explaining opportunity gaps in educational practice. *Journal of Black Studies*, *43*, 693-718. https://doi.org/10.1177/0021934712442539

National Clearinghouse for English Acquisition & Instruction Educational Programs. (2015). *Fast Facts: English Learners (Els) and NAEP*. Retrieved November 28, 2021, from https://files.eric.ed.gov/fulltext/ED564260.pdf

Nguyen, T., Watts, T. W., Duncan, G. J., Clements, D. H., Sarama, J. S., Wolfe, C., & Spitler, M. E. (2016). Which preschool mathematics competencies are most predictive of fifth grade achievement? *Early Childhood Research Quarterly*, *36*, 550-560. https://doi.org/10.1016/j.ecresq.2016.02.003

Okrent, A., & Burke, A. (2021). *Participation of Demographic Groups in STEM*. Science and Engineering Indicators. https://ncses.nsf.gov/pubs/nsb20212/participation-of-demographic-groups-in-stem

Orosco, M. J. (2014). Word problem strategy for Latino English language learners at risk for math disabilities. *Learning Disability Quarterly*, *37*, 45-53. https://doi.org/10.1177/0731948713504206

Orosco, M. J. & Abdulrahim, N. A. (2018). Examining comprehension strategy instruction with English learners' problem solving: Study findings and educator preparation implications. *Teacher Education and Special Education*, *41*, 215-228. https://doi.org/10.1177/0888406418770787

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. https://doi.org/10.1177/0022466911416248

Polat, N., Zarecky-Hodge, A., & Schreiber, J. B. (2016). Academic growth trajectories of

Multilingual learners in NAEP data: The case of fourth- and eighth-grade Multilingual learners and non-Multilingual learners on mathematics and reading tests. *The Journal of Education Research*, *109*, 541-553. https://doi.org/10.1080/00220671.2014.993461

Robertson, K. (2020). *Math instruction for English language learners*. Colorín Colorado. https://www.colorincolorado.org/article/math-instruction-english-language-learners

Rutherford, T., Farkas, G., Duncan, G., Burchinal, M., Kibrick, M., Graham, J., Richland, L., Tran, N., Schneider, S., Duran, L., & Martinez, M. E. (2014). A Randomized Trial of an Elementary School Mathematics Software Intervention: Spatial-Temporal Math. *Journal of Research on Educational Effectiveness*, *7*, 358-383. https://doi.org/10.1080/19345747.2013.856978

Samson, J. F., & Collins, B. A. (2012). *Preparing all teachers to meet the needs of English language learners: Applying research to policy and practice for teacher effectiveness*. https://files.eric.ed.gov/fulltext/ED535608.pdf

Sheng, Z., Sheng, Y., & Anderson, C. J. (2011). Dropping out of school among ELL students: Implications to Schools and teacher education. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *84*, 98-103. https://doi.org/10.1080/00098655.2010.538755

Short, D. J., Becker, H., Cloud, N., Hellman, A. B., & Levine, L. N. (2018). *The 6 Principles for Exemplary Teaching of English Learners*. Alexandria, VA: TESOL International Association.

Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities*, *49*, 77-96. https://doi.org/10.1177/0022219414528540

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (2021). *English language learners in public schools*. Retrieved May 26, 2021, from https://nces.ed.gov/programs/coe/indicator/cgf

U.S. Department of Justice & U.S. Department of Education. (2015). *Ensuring English Learner Students Can Participate Meaningfully and Equally in Educational Programs*. Office for Civil Rights. Retrieved October 20, 2022, from https://www2.ed.gov/about/offices/list/ocr/ellresources.html

Van Orman, D. S. J., Ardasheva, Y., Carbonneau, K. J., & Firestone, J. B. (2021). Examining the impacts of extended vocabulary instruction in mixed-English-proficiency science classrooms. *The Journal of Educational Research*, *114*, 74–88. https://doi.org.prox.lib.ncsu.edu/10.1080/00220671.2021.1881754

Vespa, J., Medina, L., & Armstrong, D. M. (2018). *Demographic turning points for the United States: Population projections for 2020 to 2060*. Current Population Reports. US Census Bureau. Retrieved November 15, 2021, from https://heartlandcua.org/common/Uploaded%20files/DEI/Census_projections.pdf

Wanzek, J., Vaughn, S., Scammacca, N., Metz, K., Murray, C., Roberts, G., & Danielson, L. (2013). Extensive reading interventions for students with reading difficulties after grade 3. *Review of Educational Research*, *83*, 163-195. https://doi.org/10.3102/0034654313477212.

Xin, Y. P., Kim, S. J., Lei, Q., Wei, S., Liu, B., Wang, W., Kastberg, S., Chen, Y., Yang, X., Ma, X., & Richardson, S. E. (2020). The effect of computer-assisted conceptual model-based intervention program on mathematics problem-solving performance of at-risk English learners. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, *36*, 104-123. https://doi.org/10.1080/10573569.2019.1702909

Xu, Y., & Drame, E. (2008). Culturally appropriate context: Unlocking the potential of response to intervention for English language learners. *Early Childhood Education Journal*, *35*, 305-311. https://doi.org/10.1007/s10643-007-0213-4

Yankelowitz, J. (2023, February). *A move towards multilingualism: why this term is better than others [for now]*. SRI International. https://multilinguallearner.org/a-move-towards-multilingual-terms/

Yu, J., Wei, X., Hall, T. E., Oehlkers, A., Ferguson, K., Robinson, K. H., & Blackorby, J. (2021). Findings from a Two-Year Effectiveness Trial of the Science Notebook in a Universal Design for Learning Environment. *Frontiers in Education*, *6*, 1-14. https://doi.org/10.3389/feduc.2021.719672

Zheng, B., Warschauer, M., Hwang, J. K., & Collins, P. (2014). Laptop Use, Interactive Science Software, and Science Learning among At-Risk Students. *Journal of Science Education*

Table 1.1. Characteristics of Science Studies.

Authors	Year	Grade	Intervention Type	Intervention Administrator	Group Size	Hours	Responsive Instruction	Native Language
Aguirre-Muñoz et al.	2021	K-2nd	Adult-directed + peer-mediated	Class teacher	Whole group		OD GM IE RE	Spanish
August et al.	2010	6th	Adult-directed + peer-mediated	Class teacher	Whole group	30	LO OD GM IE	Spanish
Bravo & Cervetti	2014	4th-5th	Adult-directed	Class teacher	Individual	40	LO GM	Spanish
Cervetti et al.	2015	4th-5th	Adult-directed	Class teacher	Whole group		OD NL	Spanish
Yu et al.	2021	4th	Computer-based	Class teacher	Individual		NL	Vietnamese Arabic
Zheng et al.	2014	5th	Computer-based	Class teacher	Individual		GM	Spanish

Note. English language proficiency level was not included in the table due to lack of information reported. K = kindergarten, OD = facilitating oral discussion with students, GM = using graphic organizers and manipulatives, IE = explicitly using students' own ideas and experiences, RE = providing relevant instructional examples to participants' daily lives, pop culture, and cultural heritage, NL = allowing use of native language, LO = explicitly stating measurable lesson objectives.

Table 1.2. Effect Sizes of Science Studies for Each Outcome Measure.

Authors	Year	Grade	*n*	Outcome	Measures	Effect Size
Aguirre-Muñoz et al.	2021	K-2nd	348	Science, technology, & engineering knowledge	SELA	*d* = 1.57 (K) *d* = 0.82 (1st) *d* = 0.68 (2nd)
August et al.	2010	6th	562	Science achievement	RD test	*d* = 0.25
Bravo & Cervetti	2014	4th-5th	115	Science achievement	RD tests	*d* = 0.80
Cervetti et al.	2015	4th-5th	147	Science achievement	RD tests	*d* = -0.33
Yu et al.	2021	4th	134	Science achievement	MAP test of science[a]	*d* = 0.27
Zheng et al.	2014	5th	227	Science achievement	CST science test[a]	*d* = 0.59

Note. *n* = number of participants in analytic sample, K = kindergarten, SELA = Science and Engineering Learning Assessment, RD = researcher-developed, STAAR = State of Texas Assessments of Academic Readiness, VSS-ES = Vocabulary-of-science scale-earth science, MAP = Measures of Academic Progress.
[a] = criterion- or norm-referenced measure.

Table 2.1. Characteristics of Math Group Design Studies.

Authors	Year	Grade	Intervention Type	Intervention Administrator	Group Size	Hours	Responsive Instruction	Native Language	English Proficiency, *M*; [range]
Crawford	2013	7th-8th	Computer-based	Class teacher	Individual	18	NL GM	Spanish	[1-5]
Doabler et al.	2019	K	Adult-directed	Interventionist	Small group	16.67		Spanish Arabic Portuguese	1.8; [1-4]
Freeman	2012	9th-10th	Computer-based	Class teacher	Individual	15[a]	NL GM	Spanish	
Orosco & Abdulrahim	2018	3rd	Adult-directed	Researcher-trained individuals	Small group	60		Spanish	3
Rutherford et al.	2014	K-5th	Computer-based	Class teacher	Individual	51		Spanish Vietnamese	

Note. English proficiency was coded based on WIDA and CELDT English language proficiency levels. K = kindergarten, NL = allowing use of native language, GM = using graphic organizers and manipulatives.
[a] = minimum number of hours reported.

Table 2.2. Effect Sizes of Math Group Design Studies for Each Outcome Measure.

Authors	Year	Grade	*n*	Outcome	Measures	Effect Size
Crawford	2013	7th-8th	396	Math achievement	RD test	*d* = 0.94
Doabler et al.	2019	K	287	Early numeracy[a]	RAENS Oral Counting[b] ASPENS[b] NSB[b] TEMA-3[b]	*d* = 1.19 *d* = 0.22 *d* = 0.95 *d* = 0.50 *d* = 0.31
Freeman	2012	9th-10th	50	Math achievement	Third-party test	*d* = 0.46
Orosco & Abdulrahim	2018	3rd	78	Word problem solving[a]	TOMA[b] KM[b]	*d* = 0.67 *d* = -0.05
Rutherford et al.	2014	K-5th	8,673	Math achievement	Math CST[b]	*d* = 0.06

Note. *n* = number of participants in analytic sample, RD = researcher-developed, K = kindergarten, RAENS = ROOTS Assessment of Early Numeracy Skills, ASPENS = Assessing Students Proficiency in Early Number Sense, NSB = Number Sense Brief Screener, TEMA-3 = Test of Early Mathematics Assessment – Third Edition, TOMA = Test of Mathematical Abilities, KM = KeyMath, CST = California Standards Test.
[a]This outcome has multiple dependent effect sizes that were averaged to address multiplicity.
[b] = criterion or norm-referenced measure.

Table 3.1. Characteristics of Math Single-Subject Design Studies.

Authors	Year	Grade	Intervention Type	Intervention Administrator	Group Size	Hours	Responsive Instruction	Native Language	English Proficiency, *M*; [range]
Kim et al.	2015	4th	Adult-directed	Experimenter	Individual		RE	Chinese Korean	
Kong & Swanson	2019	3rd	Adult-directed	Researcher-trained individuals	Small group	10.5		Spanish Indonesian	2; [1-3]
Luevano & Collins	2020	2nd	Adult-directed	Experimenter + Researcher-trained individuals	Small group + Individual	4.92	NL RE	Spanish	3.8; [3-4]
Orosco	2014	3rd	Adult-directed	Class teacher	Small group + Individual	6.38	OD	Spanish	
Orosco et al.	2013	2nd	Adult-directed	Experimenter + Interventionist	Individual	6.38		Spanish	[1-2]
Xin et al.	2020	3rd	Computer-based	Interventionist	Individual	11.67	GM	Spanish	

Note. English proficiency was coded based on WIDA and CELDT English language proficiency levels. RE = providing relevant instructional examples to participants' daily lives, pop culture, and cultural heritage, NL = allowing use of native language, OD = facilitating oral discussion with students, GM = using graphic organizers and manipulatives.

Table 3.2. Effect Sizes of Math Single-Subject Design Studies for Each Outcome Measure.

Authors	Year	Grade	*n*	Outcome	Measures	Effect Size
Kim et al.	2015	4th	3	Fraction word problem solving	RD items	Tau-U = 1.00 (skill 1) Tau-U = 1.00 (skill 2)
Kong & Swanson	2019	3rd	9	Word problem solving	RD items M-CAP[a]	Tau-U = 0.53 Tau-U = 0.66
Luevano & Collins	2020	2nd	4	Word problem solving	RD items	Tau-U = 0.66
Orosco	2014	3rd	6	Word problem solving	RD items	Tau-U = 0.66
Orosco et al.	2013	2nd	6	Word problem solving	RD items	Tau-U = 0.69
Xin et al.	2020	3rd	4	Word problem solving	AWPS	Tau-U = 0.96

Note. *n* = number of participants in analytic sample, RD = researcher-developed, M-CAP = AIMSweb Math Concepts and Applications, AWPS = Additive word problem solving criterion test.
[a] = norm-referenced measure.

APPENDICES

Appendix A

Code Book

Subject Area – Study includes the subject area being targeted.

Science – Study is targeting science/engineering performance.

Math – Study is targeting math performance.

Grade-Level – Study includes the grade-level of its participants.

Lower Elementary – Study participants are in grades K-2.

Upper Elementary – Study participants are in grades 3-5.

Middle School – Study participants are in grades 6-8.

High School – Study participants are in grades 9-12.

Socioeconomic Status (SES) – Study includes a measure of participants' SES.

Low – Participants are described as coming from a low-income background or demonstrate low-income/education (e.g., qualify for free or reduced school lunch).

Middle – Participants are described as coming from a middle-class background or demonstrate middle class income/education (e.g., 30% above and below the median income).

High – participants are described as coming from an upper-class background or demonstrate upper class income/education (e.g., double the median income).

Intervention Type – Study includes a description of how the intervention was delivered.

Adult-Directed – Intervention is delivered by a professional.

Peer-Mediated – Intervention is delivered through peer work.

Computer-Based – Intervention is delivered through a computer.

Intervention Administrator – If adult-directed, study includes what type of professional delivered the intervention.

Classroom Teacher – The intervention was delivered by a general education classroom teacher.

Trained Interventionist – The intervention was delivered by a school interventionist or other professional trained to deliver the intervention.

Experimenter – The intervention was delivered by a professional conducting research.

Researcher-trained individuals – The experimenter trained other individuals to assist with delivering the intervention.

Tutor – The intervention was delivered by a tutor

Group Size – The study includes how many students received the intervention at once.

Individual – The intervention was delivered one-on-one.

Small Group – The intervention was delivered in a group of no more than five students.

Medium Group – The intervention was delivered in a group of 6-15 students.

Whole Group – The intervention as delivered to all students in the classroom.

Dosage – Study includes the total number of hours that participants received the intervention or total number of hours can be calculated.

E.g., Minutes per day for an entire school year can be calculated based on an average of 180 days of instruction in a school year.

E.g., Three hours bi-weekly for an entire school year can be calculated based on an average of 36 weeks of instruction in a school year.

Effect Size – Study includes an effect size or gives the type of analysis that was run.

E.g., Effect Size – partial eta squared (0.79)

E.g., Type of Analysis – independent samples *t* test, chi-squared test of independence

Language – Study includes the language/language family spoken by its participants.

E.g., Chinese – Mandarin, Taiwanese, Cantonese, and Hmong

Level of Language Development – Study includes a measure of participants' level of language development.

WIDA or CELDT Level 1 (Entering or Beginning) – Study participants are described as having English proficiency on WIDA or CELDT Level 1, WIDA "entering," or CELDT "beginning."

WIDA or CELDT Level 2 (Beginning or Early Intermediate) – Study participants are described as having English proficiency on WIDA or CELDT Level 2, WIDA "beginning," or CELDT "early intermediate."

WIDA or CELDT Level 3 (Developing or Intermediate) – Study participants are described as having English proficiency on WIDA or CELDT Level 3, WIDA "developing," or CELDT "intermediate."

WIDA or CELDT Level 4 (Expanding or Early Advanced) – Study participants are described as having English proficiency on WIDA or CELDT Level 4, WIDA "expanding," or CELDT "early advanced."

WIDA or CELDT Level 5 (Bridging or Advanced) – Study participants are described as having English proficiency on WIDA or CELDT Level 5, WIDA "bridging," or CELDT "advanced."

Mixed – Study participants are described as having different English proficiency levels (e.g., "entering to developing" or Levels 1-4). Note the average WIDA or CELDT level if reported. If results are given by level of language development, note each effect size. Or note a significant interaction effect.

Culturally and Linguistically Responsive Instruction (Driver and Powell, 2017) – Study includes how culturally and linguistically responsive instruction was incorporated into the intervention.

E,g., Explicitly stating measurable lesson objectives, facilitating oral discussions with students, allowing use of native language, using graphic organizers and manipulatives, explicitly using students' own ideas and experiences, providing relevant instructional examples to participants' daily lives, pop culture, and cultural heritage.