A Meta-Analysis of Single-Case Research on Word-Problem Instruction for Students With Learning Disabilities

Mikyung Shin, PhD[1] [iD], Diane P. Bryant, PhD[2], Sarah R. Powell, PhD[2], Pyung-Gang Jung, PhD[3], Min Wook Ok, PhD[4], and Fangjuan Hou, PhD[5]

Abstract

In this meta-analysis of single-case designs, the researchers investigated the effects of word-problem instruction for students with learning disabilities in Grades 1 through 12, as reported in 20 peer-reviewed articles published between 1975 and 2020. A two-level multilevel model with a robust variance estimation method was implemented to calculate the effect sizes. A significant overall effect of word-problem instruction for students with learning disabilities was found, and the results indicate that studies meeting all three quality indicators of implementation fidelity (i.e., adherence, dosage, and regularity) showed greater effects. The average weighted effect sizes of word-problem instruction varied depending on the content and practice standards of the Common Core State Standards for Mathematics. Limitations and future research directions are also discussed.

Keywords

learning disabilities, meta-analysis, moderator analysis, single-case design, word-problem instruction

The development of mathematical proficiency in solving word problems (WPs) is a critical component of mathematics education for all students at the elementary and secondary levels. As a result, instruction about WPs must occur early on as part of formal education for all students. Notably, instruction on WP proficiency occurs as early as kindergarten with simple addition and subtraction WPs, and some of these tasks may be presented orally without requiring students to read (National Governors Association Center for Best Practices & Council of Chief State School Officers [NGA Center & CCSSO], 2010). Students' proficiency with WPs requires increased attention throughout the grades as they learn about more complex types of WPs (e.g., two- or three-step WPs) and WPs that capture more advanced mathematics skills (e.g., rational numbers and algebra).

Mathematical proficiency with WPs is further enhanced by the integration of mathematical practices (MPs; for example, reasoning quantitatively and attending to the structure of problems) stemming from the Common Core State Standards for Mathematics (CCSSM; NGA Center & CCSSO, 2010). For instance, WPs involve certain structures (e.g., compare, equal groups) that, when recognized and understood by students, can facilitate successful WP solutions. Mayer

(1998) indicated, however, that WPs can be challenging because students must be able to transform linguistic (syntax) and schematic (problem structure) content into a quantitative, graphic, or symbolic representation and then apply a strategy to solve the problem (Montague et al., 2011; National Mathematics Advisory Panel, 2008; Powell, 2011). Finding a solution might entail applying a meta-cognitive strategy that specifies steps in the WP process and includes a self-regulation aspect, such as asking oneself questions to determine the problem and extract relevant information (Pfannenstiel et al., 2015). These WP concepts and skills are essential for students at the elementary and secondary levels, particularly those who experience difficulty with mathematics.

[1]West Texas A&M University, Amarillo, TX, USA
[2]The University of Texas at Austin, Austin, TX, USA
[3]Korea National University of Education, Cheongju-si, Republic of Korea
[4]Daegu University, Gyeongsan-si, Republic of Korea
[5]Independent scholar

Corresponding Author:
Mikyung Shin, Department of Education, West Texas A&M University, Amarillo Center 332D, 720 S. Tyler, Amarillo, TX 79101, USA.
Email: mshin@wtamu.edu

WP Difficulties of Students With Learning Disabilities

Students with a learning disability (LD) often demonstrate lower performance on assessments of WPs than students without LDs (Wei et al., 2013), and lower WP performance begins in early elementary grades. For example, when Jordan and Hanich (2000) administered addition and subtraction WPs to second-grade students, students with LDs answered fewer WPs correctly than students without LDs and also employed less efficient WP strategies, such as automatic processing and retrieval. Vukovic et al. (2010) reported findings similar to those of Jordan and Hanich (2000) with third-grade students. Andersson (2008) determined performance differences in the third and fourth grades on one-and multi-step WPs; students without LDs exceeded performance levels of students with LDs on both types of WP assessments. Fuchs and Fuchs (2002) administered a measure of complex WPs comprising four types (i.e., shopping list, buying bags, half, and pictographs) and multi-step WPs with tables and graphs to fourth graders. In both WP sets, students with LDs scored significantly lower those without LDs. Kingsdorf and Krawec (2014) examined the performance of seventh and eighth graders on one-, two-, and three-step WPs. In an analysis of six common problem-solving errors (i.e., number selection, operation, missing, computation, random, and omission), students with LDs demonstrated more errors than those without LDs.

This brief review demonstrates that solving WPs is an area of concern for students with LDs. Students with LDs often experience significant difficulties with WPs due to the lack of foundational skills necessary for successful WP solving. First, solving WPs requires not only mathematics but also reading competency (Boonen et al., 2016). When solving WPs, the first thing students should do is to read and understand the WP's text; however, many students with LDs experience difficulty with reading and reading comprehension (Peake et al., 2015). Students with LDs often misunderstand or fail to conceptualize the semantic WP structures, and the deficits may compound their difficulties with WPs (Cortiella & Horowitz, 2014). Second, students with LDs often demonstrate difficulty translating information from within the WP prompt into equations or visuals because of their limited abstract reasoning skills (Kingsdorf & Krawec, 2014). They may have trouble determining whether to add, subtract, multiply, or divide. After a problem is translated into an equation, students with LDs also have challenges in applying relevant arithmetic strategies (e.g., recall basic facts and computation), and this can be another major obstacle to successfully solving WPs (Bryant et al., 2000; Tolar et al., 2016). Finally, WPs are difficult for students with LDs because of the complex nature of solving WPs. To solve a WP, students need to organize information and undergo numerous problem-solving steps (van Garderen et al., 2014). However, this can challenge students with LDs who often have difficulty memorizing and correctly following the WP-solving steps because of working memory (Shin & Bryant, 2015b).

Previous Syntheses and Meta-Analyses of WP Instruction

Due to the need to help students with LDs learn to solve WPs efficiently and effectively, research in special education has led to the development of interventions focused on improving WP outcomes. Thus, over the last three decades, researchers have examined the effectiveness of WP instruction for students with LDs and mathematics difficulties (S. C. Cook et al., 2020; Hwang & Riccomini, 2016; Jitendra et al., 2015, 2016, 2018; Jitendra & Xin, 1997; Montague & Dietz, 2009; Peltier et al., 2018; Powell, 2011; Xin & Jitendra, 1999; Zheng et al., 2013). These studies compared the effects of WP instruction by student characteristics such as disability status (e.g., LDs and mathematics difficulties) or grade (e.g., elementary, secondary, and postsecondary) and intervention features such as an interventionist (e.g., researcher, teacher, and both), WP tasks (e.g., arithmetic or fractions, ratio and proportion, and algebra), and instructional components (e.g., use of visuals, manipulatives, technology, and strategy).

In general, students with LDs and mathematics difficulties demonstrated improved performance on WP tasks after receiving WP instruction, and the weighted average effect sizes showed significant improvements. Many researchers (S. C. Cook et al., 2020; Hwang & Riccomini, 2016; Jitendra et al., 2015; Jitendra & Xin, 1997; Powell, 2011; Zheng et al., 2013) identified positive effects from the use of explicit and systematic instruction, in terms of modeling, guided practice, monitoring, corrective feedback, or independent practice. Other researchers (e.g., S. C. Cook et al., 2020; Jitendra et al., 2015, 2016, 2018; Peltier et al., 2018; Powell, 2011; Xin & Jitendra, 1999) determined a high degree of effectiveness when incorporating representational techniques, including the use of schematic diagramming with the identification of WP structures. Furthermore, researchers identified the use of heuristic strategy, a general WP-solving approach (e.g., read the problem, highlight keywords, solve the problem, and check your work; Gersten et al., 2009), as another effective instructional component (Hwang & Riccomini, 2016; Jitendra et al., 2015, 2018; Montague et al., 2011).

The combination of varied instructional components, including visual models combined with other strategies such as the heuristic strategies, yielded higher effects compared with single-component intervention for students with LDs and mathematics difficulties (Hwang & Riccomini, 2016; Jitendra et al., 2015; Jitendra & Xin, 1997). These

findings demonstrated that researchers have continuously focused on various instructional practices when designing WP instruction to improve students' WP solving.

Assessment of Study Quality

With the increased interest in establishing the rigor of studies' quality in social science fields and investigating evidence-based practices, researchers and organizations (e.g., Council for Exceptional Children [CEC], 2014; Gersten et al., 2005; Horner et al., 2005; What Works Clearinghouse [WWC], 2011) have proposed guidelines for assessing evidence of the quality of research studies. These guidelines are known as quality indicators (QIs) that can be implemented to judge the rigor of studies. Reflecting the significance of evaluating QIs in the systematic review process, researchers (S. C. Cook et al., 2020; Jitendra et al., 2015, 2016; Montague & Dietz, 2009) examined QIs when reviewing studies on WP instruction for students with LDs and mathematics difficulties. Specifically, Montague and Dietz (2009) evaluated the QIs of studies of cognitive strategy instruction on WP instruction, employing the criteria suggested by Horner et al. (2005) for single-case design (SCD) studies and Gersten et al. (2005) for group design studies. In a follow-up study, Jitendra et al. evaluated the QIs of studies on strategy instruction and representation of problems in 2015 and 2016, respectively. They also employed assessment criteria suggested by Gersten et al. (2005) for group studies; in addition, Jitendra et al. (2015) used Horner et al. (2005) and the WWC (2011) for SCD studies. Most recently, S. C. Cook et al. (2020) evaluated QIs for studies of schema-based instruction (SBI) using the CEC's (2014) QI criteria.

Overall, with the implementation of QIs, Montague and Dietz (2009) did not support cognitive strategy instruction as an evidence-based practice due to a lack of rigor of the included studies. Notably, researchers who conducted and published studies prior to 2005 did not have the benefit of designing interventions to the rigorous level expected in research papers. In addition, S. C. Cook et al. (2020) suggested that SBI might be an evidence-based practice, and Jitendra et al. (2015, 2016) determined that strategy instruction with the use of visual representations might constitute an evidence-based practice for students with LDs and mathematics difficulties.

CCSSM

Since the release of mathematics learning expectations (CCSSM) for elementary and secondary students in 2010, mathematical proficiency with WPs has been enhanced with the integration of these standards at schools (NGA Center & CCSSO, 2010). Specifically, the CCSSM, which acknowledges problem-solving skills as one of the essential skills of the 21st century, specifies solving WPs as the primary area of focus for all students, including those with disabilities (National Council of Teachers of Mathematics, 2014). Thus, across the CCSSM content (e.g., operations and algebraic thinking [OA], number and operations—fractions [NF], and expressions and equations [EE]) and practice standards (e.g., making sense of problems, reasoning abstractly and quantitatively, and modeling with mathematics), the standards emphasize WP opportunities and practices.

Some researchers have evaluated potential frameworks regarding how WP instruction could be aligned with CCSSM standards (Shin & Bryant, 2015a; Spooner et al., 2017). Shin and Bryant (2015a) reported that 11 of the 17 studies meeting the NF CCSSM content standard focused on WPs, and the most frequently observed CCSSM practice standard was modeling with mathematics. Spooner et al. (2017) also stated that when teaching WPs to students with intellectual disabilities, identifying and making sense of problem-solving, analyzing quantitative relationships in mathematical contexts, and modeling with mathematics could be used. However, due to limited reviews on the application of CCSSM standards as their main focus of data analysis, additional studies are needed to ensure the validity of the analysis of mathematics topics and practices. Although these previous studies (Shin & Bryant, 2015a; Spooner et al., 2017) have not examined the moderation effects of CCSSM standards, several previous meta-analyses (Jitendra et al., 2016, 2018; Lein et al., 2020) suggested that CCSSM content (i.e., mathematical topics) and practice standards (i.e., instructional approach) could be potential moderators affecting the effects of mathematics interventions of WPs for students experiencing difficulty with mathematics, including those having LDs.

The Need for This Meta-Analysis

The present meta-analysis updates and extends previous syntheses and meta-analyses on teaching WPs disaggregating data on students with LDs (S. C. Cook et al., 2020; Jitendra et al., 2016; Lein et al., 2020; Peltier et al., 2018; Xin & Jitendra, 1999). Although four previous syntheses (S. C. Cook et al., 2020; Jitendra et al., 2015, 2016; Montague & Dietz, 2009) attempted to evaluate the QIs of studies on WP instruction for students with disabilities and mathematics difficulties, they mainly focused on specific types of interventional practices—cognitive strategy instruction (Montague & Dietz, 2009), use of representational models (Jitendra et al., 2016), strategy instruction (Jitendra et al., 2015), and SBI (S. C. Cook et al., 2020). Therefore, there is a need to appraise the quality of studies applying the CEC (2014) QIs to broad aspects of WP interventions beyond these previously evaluated intervention types, particularly targeting the focus group (students with LDs) in the current

meta-analysis. Furthermore, Jitendra et al. (2015), as well as Montague and Dietz (2009), applied the QI criteria by Horner et al. (2005) or the WWC (2011). More recently, S. C. Cook et al. (2020) implemented the QI criteria by the CEC (2014) to evaluate the study quality (CEC, 2014) of WP instruction. None of these studies quantifiably examined the effect of study quality as a moderating variable. Consequently, there is a clear need to evaluate the moderating effect of study quality to validate findings from previous syntheses.

Furthermore, in previous meta-analyses (Jitendra et al., 2016, 2018; Lein et al., 2020), researchers indicated that mathematical topics and instructional approaches, which were aligned with CCSSM content and practice standards, could influence the effects of WP interventions for students struggling with mathematics, including those with LDs. However, neither of the previous meta-analyses (Jitendra et al., 2018; Peltier et al., 2018) specifically targeted the CCSSM content and practice standards (NGA Center & CCSSO, 2010) for their main moderating variables beyond problem types (Peltier et al., 2018) or intervention approach (Jitendra et al., 2018). Considering the increasing educational interest and expectations aligned with CCSSM, it is necessary to further investigate the effects of WP instruction following CCSSM standards. Furthermore, by controlling for covariates within CCSSM standard categories, we can examine the moderating effects of each CCSSM content and practice standard.

In particular, we focused on SCD studies in which researchers implemented experimental designs and documented the functional relationships between independent and dependent variables (Horner et al., 2005). Considering the nested structure of SCD studies (measurement occasions were nested within each case, and cases were nested within each study), we applied a two-level multilevel model to examine the effect sizes between cases (Baek et al., 2016; Moeyaert et al., 2020; Pustejovsky et al., 2014). Considering the dependency on multiple outcome measures within a study, we also applied a robust variance estimation (RVE) method to combine effect sizes across studies and estimated standard errors and confidence intervals (CIs) around the average intervention effect estimates (Hedges et al., 2010; Tipton, 2015; Tipton & Pustejovsky, 2015).

Given the aforementioned need for an analysis of WP studies implementing SCD studies, the purpose of the current meta-analysis was to investigate the overall effect of WP instruction and the moderating effects of study quality based on the CEC (2014) QIs and CCSSM content and practice standards (NGA Center & CCSSO, 2010) for students with LDs in Grades 1 through 12. Specific research questions to meet the study purpose were as follows:

Research Question 1 (RQ1): What is the overall effect of WP instruction for students with LDs?

Research Question 2 (RQ2): To what extent do the studies meet the CEC QIs, and does the study quality moderate the overall effect?
Research Question 3 (RQ3): To what extent do studies meet the CCSSM content standard, and does the CCSSM content standard moderate the overall effect?
Research Question 4 (RQ4): To what extent do studies meet the CCSSM practice standard, and does the CCSSM practice standard moderate the overall effect?

Method

To maintain consistency and ensure standardization in reporting the methodology of the current meta-analysis, we followed the Meta-Analysis Reporting Standards for quantitative research syntheses (Appelbaum et al., 2018).

Article Selection Procedures

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for article selection procedures (Moher et al., 2009). The focus of the current meta-analysis was an examination of the effects of WP-solving instruction for students with LDs published between 1975 and 2020. The year 1975 was when the category of LDs was included in the Education for All Handicapped Children Act (Pub. L. No. 94-142). Figure S1 in the Supplemental Material presents the adapted flow diagram.

Inclusion criteria. Studies were included in this meta-analysis based on the following inclusion criteria: (a) participants were students with LDs who were identified by school districts as having LDs and were in Grades 1 through 12; (b) WP-solving instruction was the independent variable, targeting the teaching of WP solving; (c) the research design should be SCD, either multiple-baseline or multiple-probe across subjects, AB^k (k = number of phase contrasts) withdrawal designs targeting, or alternating treatment designs with baseline; and (d) there should be at least three participants in each study, and a minimum of three data points per baseline and intervention phase are needed to calculate between-case effect sizes using a two-level multilevel model.

Search strategies. We conducted an electronic database search of *Education Resources Information Center* (*ERIC; n* = 1,181), *PsycINFO* (*n* = 1,982), and *Academic Search Complete* (*n* = 1,899) for peer-reviewed articles published between 1975 and 2020. Keywords included the following terms: (learning disab* OR dyslexi* OR dyscal* OR math* disab*) AND (word problem* OR problem solving*). We identified 5,062 articles with this initial search: 1,620 duplicated articles were removed, resulting in 3,442 articles.

After reviewing the titles and abstracts of 3,442 articles, we excluded 3,097 articles. Full-text copies of the remaining 345 articles were retrieved. After two authors reviewed the 345 articles, using the four inclusion criteria (mentioned previously), we excluded 325 articles for the following reasons: (a) participants were not students with LDs ($n = 33$); (b) independent variables were not WP-solving instructions or outcome measures were not related to WP solving ($n = 149$); (c) the research designs were randomized controlled trials, quasi-experimental design, or SCDs with AB[k] without baseline or alternating treatment without baseline or multiple-baseline/multiple-probe across behaviors or settings ($n = 140$); and (d) fewer than three participants with fewer than three data points per phase ($n = 3$). As a result, 20 articles that met all the inclusion criteria were identified. As the last step, we conducted an ancestral search of the reference lists of the 20 identified articles. We did not find any additional articles through this inverse search. Two authors checked the interrater reliability of the screening of studies; 100% interrater reliability was achieved using a percent agreement formula, taking the number of agreements divided by the sum of agreements and disagreements and multiplied by 100% (Araujo & Born, 1985, p. 208). Appendix A in the Online Supplemental Material reports the lists of included studies.

Coding Procedures

Coder qualification and training. Four main coders with at least 10 years of experience (e.g., systematic review and meta-analysis) in the field of special education participated in the coding process. Before coding each article, the first author provided 1-hr face-to-face coder training and three to four follow-up individual checkups via email and phone on the definitions of coding variables and on using a coding spreadsheet for code data. After the training, coders independently coded one article and achieved an interrater reliability of 97%; they met 100% agreement after discussion. For coder training in extracting SCD data, the first author provided an additional one-time training of approximately 30 min. The coders then independently coded one multiple-baseline across subjects and attained an interrater reliability of 98%; after checking the graphed data together, they reached 100% agreement.

Study characteristics. To provide an overall summary of study characteristics, we coded 20 identified articles using a researcher-developed Excel spreadsheet coding protocol. Study characteristics included the following variables: participants' demographic information such as the number of cases, age, and gender; area of difficulty; type of research design (e.g., multiple-baseline or multiple-probe); type of interventionist (e.g., researcher or special education

teacher); type of intervention (e.g., self-regulated strategy, SBI); and length of instructional time per session.

Assessment of study quality. The first moderator we examined was study quality. The evaluation of study quality helps authors monitor any risk of bias within each study and facilitates the transparency of their review process (Viswanathan et al., 2018). To examine the methodological quality of the included 20 studies, we applied the CEC (2014) QIs. The CEC suggested 22 QIs across eight categories for SCD studies: 1.0. context and setting (1.1. critical features of the context or setting), 2.0. participants (2.1. participant demographics and 2.2. disability status and method of the participants), 3.0. intervention agents (3.1. intervention agent's role and 3.2. specific training), 4.0. descriptions of practice (4.1. intervention procedures and 4.2. intervention materials), 5.0. implementation fidelity (5.1. adherence, 5.2. dosage, and 5.3. regularity), 6.0. internal validity (6.1. systematical manipulation of the independent variable, 6.2. description of the baseline, 6.3. no access to the intervention during baseline, 6.5. at least three demonstrations of the experimental effects, 6.6. at least three data points during the baseline, and 6.7. control common threats to internal validity), 7.0. outcome measures and dependent variables (7.1. socially important, 7.2. description of the outcome measures, 7.3. graphed data, 7.4. frequency and timing of the outcome measures, and 7.5. adequate interrater reliability), and 8.0. data analyses (8.2. clear graphs representing the outcome data). All studies were reported as meeting multiple CEC QIs and were coded for each applicable QI. Each QI was rated "1" when meeting the criterion when the included study satisfactorily addressed the concern, and "0" when not meeting the criterion due to a lack of reported information. Therefore, each study could receive a total score ranging from 0 to 22.

CCSSM content standard. We coded the CCSSM content standard (NGA Center & CCSSO, 2010) as the second moderator. Based on the definitions of CCSSM content standards, we first examined whether a given study targeted each of the WP concepts and skills in Grades 1 to 12. Six domains were observed across studies: OA, NF, geometry (G), ratios and proportional relationships (RP), the number system (NS), and EE. Studies were coded "1" when meeting a particular domain standard and "0" when not meeting the standard.

CCSSM practice standard. We also coded the CCSSM practice standard (NGA Center & CCSSO, 2010) as the third moderator. According to the definitions and examples of CCSSM practice standards, we analyzed all the descriptions regarding interventions, materials, and measures in each study. We identified and coded data in the following eight domains: *make sense of problems and persevere in*

solving them (MP 1), *reason abstractly and quantitatively* (MP 2), *construct viable arguments and critique the reasoning of others* (MP 3), *model with mathematics* (MP 4), *use appropriate tools strategically* (MP 5), *attend to precision* (MP 6), *look for and make use of structure* (MP 7), and *look for and express regularity in repeated reasoning* (MP 8). All studies were reported to meet multiple CCSSM practice standards and were coded for each applicable category. Studies were coded "1" when meeting a particular practice standard and "0" when not meeting the standard.

Extraction of SCD data. The GetData Graph Digitizer 2.26 web-based application (GetData Graph Digitizer, 2013) was used to extract baseline and intervention data from the line graphs of each study. To collect visual data, researchers conducted the following procedures: (a) opened an image graph from the desktop folder, (b) set the minimum and maximum values of the *x*- and *y*-axes to compare data within each graph, (c) clicked the data in each baseline or intervention phase and transformed them into raw numerical data, and (d) copied and pasted the *y*-axis values, which are WPs (dependent variable), to the Excel file. The reliability ($r = 1.00$, $p < .001$) and validity ($r = 1.00$, $p < .001$) for the software were significantly high (Shin & Jung, 2018).

Interrater reliability for article coding. The first author coded all articles, and the other three authors independently coded and checked all 20 articles to identify any disagreements on each variable. We calculated Cohen's kappa coefficient (κ) for interrater reliability. According to Altman's (1999) guidelines, there were strong agreements between the first coder and the others: κ was .98 for study characteristics, .98 for assessment of study quality, .92 for CCSSM content standard, .96 for CCSSM practice standard, and .97 for extracted data ($ps < .001$). After discussing any disagreed data, the coders reached 100% agreement on all data.

Data Analysis

Effect size calculation. To examine the overall between-case effect size for each study, we applied a two-level (within-case changes at Level 1 and between-case variances at Level 2) multilevel model. First, the multilevel model was estimated using the lme function from the nlme R package (Pinheiro et al., 2020). Previous studies of the multilevel meta-analysis of SCD studies (Baek et al., 2016; Moeyaert et al., 2020; Pustejovsky et al., 2014) suggested a piecewise linear regression model, noting a breakpoint at the value of sessions (or time) where the level of students' WP performance is shifted. We set up the equation at Level 1 as follows: $y_{ij} = \beta_{0j} + \beta_{1j} Treatment_{ij} + \beta_{2j} Time_{ij} \times Treatment_{ij} + e_{ij}$ and $e_{ij} \sim N(0, \Sigma e)$ for the *j*th case ($j = 1, 2, \ldots, J$), where y_{ij} = WP-solving outcome

at a measurement occasion, *i*th measurement occasion ($i = 0, 1, \ldots, I$), β_{0j} = baseline level, β_{1j} = immediate treatment effect, β_{2j} = effect of the treatment on the linear time trend, and e_{ij} = residual that indicates within-case variation, assumed to be normally distributed with a mean of zero. *Treatment$_{ij}$* was dummy coded (0 for the baseline phase, 1 for the treatment phase); *Time* was centered at the first session of the treatment. In particular, the baseline outcomes were assumed to be stable. Furthermore, the equation at Level 2 assumes variability across cases at the baseline level, treatment effect, and time-by-treatment interaction as follows:

$$\begin{cases} \beta_{0j} = \theta_{00} + u_{0j} \\ \beta_{1j} = \theta_{10} + u_{1j} \\ \beta_{2j} = \theta_{20} + u_{2j} \end{cases} \text{ with } \begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{bmatrix} \sim N(0, \Sigma e),$$

again for the *j*th case, where θ_{00} = average baseline level (intercept), θ_{10} = average increase from baseline to treatment phases, θ_{20} = average change in slope between baseline and treatment phases, and (u_{0j}, u_{1j}, u_{2j}) = a random deviation for the *j*th case, assumed to be multivariate normally distributed with a mean of zero.

Next, an adjusted restricted maximum likelihood (REML) estimation, between-case standardized mean difference (BC-SMD) was calculated using the g_REML function from the scdhlm package (Pustejovsky, 2016; Pustejovsky et al., 2014); 95% CIs, autocorrelations, and intraclass correlations were also calculated.

Meta-analysis for effect sizes. This meta-analysis applied weighted linear regression models with RVE, which were clustered by study, with small-sample corrections (Hedges et al., 2010; Tipton, 2015; Tipton & Pustejovsky, 2015). First, to calculate the overall effect size without considering study covariates, we applied a random-effects model with correlated effects weights using the robu function from the robumeta R package (Fisher et al., 2017) with $\rho = 0.80$ for the mean correlation between all possible effect sizes within a cluster (at the study level). The results of sensitivity analyses support τ^2 (between-study sampling variance) and average effect sizes as being relatively robust to different ρ values (Fisher et al., 2017). Between-study heterogeneity was evaluated using the I^2 statistic, which indicates the heterogeneity ratio to the total variance across effect sizes; $I^2 > 75\%$ was considered to indicate a large degree of statistical heterogeneity (Higgins et al., 2003).

Next, to examine whether study quality (CEC, 2014) and CCSSM standards (NGA Center & CCSSO, 2010) moderate the effects of WP-solving instruction, we implemented a mixed-effects (meta-regression) model with hierarchical effects weights (Fisher et al., 2017). This generated a robust standard error, *t* test of a hypothesis test,

and CI. Due to the existence of multicollinearity among variables and the relatively small number of included studies, we ran three meta-regressions separately (Borenstein et al., 2009) for each study quality (CEC QIs), CCSSM content standard, and CCSSM practice standard. For CCSSM standards, which are categorical moderators, we created dummy-coded variables to compare moderation effects among categories (Keith, 2019).

Publication bias. We assessed publication bias by considering the possibility that the available studies under review could be a biased sample of all possible samples (Shadish et al., 2015). To identify any potential publication bias, we first visually inspected asymmetry in a scatterplot of BC-SMDs using a funnel plot (see Figure S2 in the Supplemental Material). We then statistically tested for publication bias using Egger's regression test (Egger et al., 1997). An asymmetric plot indicated a possible presence of publication bias, and Egger's test showed significant publication bias ($t = 6.50$, $p < .001$). Rothstein et al. (2005) cautioned the likelihood of potential publication bias in a small study, such as SCD studies, when studies with high positive findings are more likely to be published. The existence of publication bias in the current meta-analysis also suggests this small-study bias and suggests careful attention when interpreting the results. Data files and R scripts are available online (Shin et al., 2020).

Descriptive analysis. To provide supplemental descriptions of the meta-analysis findings, we conducted a descriptive analysis for each quantitative data analysis procedure. Odom et al. (2018) cautioned against reporting only the average effect sizes in syntheses; it is likely that some qualitative findings and observations across studies could be underestimated in doing so. Thus, we provide additional descriptions of results to enhance the understanding of the quantitative magnitude of effect sizes and provide an in-depth interpretation across studies.

Results

Summary of Study Characteristics

The 20 studies spanned a time period of more than four decades. Overall, 70% of the studies ($n = 14$) have been published since 2000. There were 82 students with LDs in Grades 3 through 12 across the 20 studies, with an average of 4 students per study (min = 3, max = 10, median = 3, $SD = 1.83$), of whom 55 were males and 27 were females. All participants had difficulty in mathematical problem-solving; 16 students were reported to have additional difficulty in reading or writing. In most studies, researchers utilized either a multiple-probe ($n = 15$) or multiple-baseline ($n = 5$) design across subjects; one study (Dennis

et al., 2016) used an AB design with a baseline. In most studies ($n = 15$), researchers provided interventions, and there were a limited number of studies in which special education teachers ($n = 5$) provided WP-solving interventions. The studies implemented several different types of interventions. Although each intervention consisted of various instructional practices, the primary types of interventions across the 20 studies were self-regulated, mnemonic strategy ($n = 7$), model-drawing strategy ($n = 2$), schema instruction ($n = 5$), explicit inquiry-based instruction ($n = 2$), video modeling ($n = 2$), and use of physical or virtual manipulatives ($n = 2$). The instructional time ranged from 18 to 55 min. Table S1 in the Supplemental Material presents a summary of the study characteristics.

Overall Effect of WP-Solving Instruction for Students With LDs (RQ1)

Figure S3 in the Supplemental Material presents forest plots of study-specific and effect size estimates (BC-SMDs) for each study. Ten of the 20 included studies reported multiple observed effect sizes, yielding a total of 39 BC-SMDs. Additional information on 95% CIs, autocorrelation, and intraclass correlation is reported in Table S2.

Table 1 shows the results of an RVE random-effects model with correlated effects weights. The overall weighted effect size estimate of WP-solving instruction for students with LDs was 4.52 (95% CI = [3.13, 5.90], $p < .001$), demonstrating a significantly large intervention effect. As shown in Table S2, all 39 effect size estimates showed positive effects of WP instruction, indicating improvement from baseline to intervention phases (min = 0.41, max = 23.05, median = 3.05, $SD = 4.85$). Regarding the measurement error, autocorrelation within case varied between −.80 and .68 (median = −.04, $SD = .34$). The intraclass correlation, which is the proportion of total variance made up by between-case variance, also varied between less than .001 and .80 (median = .16, $SD = .28$). The between-study heterogeneity was large, with $I^2 = 81.72\%$ of the variance across studies due to heterogeneity rather than chance.

Study Quality by CEC QIs and Moderating Effect (RQ2)

The first moderator was study quality. Results from the study quality by the CEC (2014) QIs are provided in Table S3. Of the 20 studies, 3 (15.00%) met all 22 of the CEC (2014) QI criteria (a total score of 22) for SCD studies, and 18 ($n = 2$), 19 ($n = 2$), 20 ($n = 2$), and 21 ($n = 11$) QI criteria. In particular, all studies met 16 (1.1, 2.1, 2.2, 3.1, 4.1, 4.2, 5.2, 6.1, 6.3, 6.5, 6.6, 7.1, 7.2, 7.3, 7.4, and 8.2) of the 22 QI criteria. The two least frequently observed CEC QIs were any specific training or qualifications of the interventionist (QI 3.2, $n = 9$) and the regularity of implementation

Table 1. Effects of WP-Solving Instruction for Students With LDs.

Variables	K	BC-SMD	SE	95% CI	I^2
Overall effect	39	4.52***	0.66	[3.13, 5.90]	81.72
CEC quality indicator[a]					
1.0. Context and setting	39	4.52***	0.66	[3.13, 5.90]	81.72
2.0. Participants	39	4.52***	0.66	[3.13, 5.90]	81.72
3.0. Intervention agent	16	4.75**	1.13	[2.13, 7.38]	87.62
4.0. Description of practice	39	4.52***	0.66	[3.13, 5.90]	81.72
5.0. Implementation fidelity	19	5.54**	1.17	[2.81, 8.27]	76.19
6.0. Internal validity	37	4.49***	0.69	[3.03, 5.95]	82.21
7.0. Outcome measures	36	4.71***	0.75	[3.12, 6.29]	82.60
8.0. Data analysis	39	4.52***	0.66	[3.13, 5.90]	81.72
CCSSM content standard					
OA	13	2.98***	0.46	[1.82, 4.14]	52.91
NF	3	5.50	2.62	[−6.14, 17.10]	78.67
G	3	11.80	4.86	[−49.90, 73.60]	85.20
RP	2	5.07	1.75	[−17.20, 27.30]	78.38
NS	10	3.04*	0.20	[0.53, 5.55]	51.50
EE	8	6.21	2.63	[−2.16, 14.60]	93.16
CCSSM practice standard					
MP 1	36	3.98***	0.56	[2.79, 5.18]	78.50
MP 2	39	4.52***	0.66	[3.13, 5.90]	81.72
MP 3	17	4.24**	0.92	[2.09, 6.39]	74.70
MP 4	39	4.52***	0.66	[3.13, 5.90]	81.72
MP 5	39	4.52***	0.66	[3.13, 5.90]	81.72
MP 6	36	3.98***	0.56	[2.79, 5.18]	78.50
MP 7	8	6.21	2.63	[−2.16, 14.60]	93.16
MP 8	—	—	—	—	—

Note. LDs = learning disabilities; WP = word problem; BC-SMD = between-case standardized mean difference; CI = confidence interval; CEC = Council for Exceptional Children; CCSSM = Common Core State Standards for Mathematics; OA = operations and algebraic thinking; NF = number and operations—fractions; RP = ratios and proportional relationships; NS = number system; EE = expressions and equations; MP = mathematical practice; G = geometry.
[a]Each study was coded "1" when meeting all quality indicators for each category (see the "Method" section).
*$p < .05$. **$p < .01$. ***$p < .001$.

fidelity across interventionists, settings, participants, or type of fidelity assessed (QI 5.3, $n = 10$), followed by adherence to implementation fidelity with an observation checklist of critical instructional elements (QI 5.1, $n = 16$), adequate interrater reliability (QI 7.5, $n = 18$), description of baseline (QI 6.2, $n = 19$), and control of common threats to internal validity (QI 6.7, $n = 19$).

Furthermore, as shown in Table 1, the 22 CEC QI criteria were categorized under eight aspects. Each weighted average effect size estimate (BC-SMD) of WP instruction for students with LDs was large and statistically significant for all eight categories, context and setting, participants, intervention agent, description of practice, implementation fidelity, internal validity, outcome measures, and data analysis, ranging from 4.52 to 5.54, $ps < .01$. All the between-study heterogeneity (I^2) for each QI category was large (min = 76.19%, max = 87.62%, median = 81.72%, $SD = 3.07$).

Table 2 shows the results of an RVE mixed-effects (meta-regression) model with hierarchical effects weights.

At the average study quality score of 20.56, there was a significant effect of WP-solving instruction from the baseline to the intervention phases ($\beta = 3.95$, $p < .001$). When the study quality score increased, the weighted effect size estimate increased slightly; however, the relationship between study quality and the magnitude of effect size estimates was not statistically significant ($\beta = 0.08$, $p = 1.34$).

Extent of Meeting CCSSM Content Standards and Moderating Effect (RQ3)

The second moderator was the CCSSM content standard (NGA Center & CCSSO, 2010). The results for the CCSSM content standards are provided in Table S4. The most frequently implemented CCSSM content standard in teaching WP solving for students with LDs was OA ($n = 8$) in solving one- and two-step four operations of whole numbers (2.OA.A.1, 3.OA.A.3, 3.OA.D.8), followed by the EE ($n = 4$)

Table 2. Meta-Regression Results for Examining WP-Solving Instruction.

Variables	Estimate (β)	SE	t	95% CI	p value
Study quality					
Intercept	3.95***	0.67	5.91	[2.44, 5.46]	<.001
CEC quality indicator[a]	0.08	0.49	0.17	[−1.17, 1.34]	1.34
CCSSM content standard[b]					
NF	5.14	2.34	2.20	[−5.74, 16.00]	.17
G	12.21	4.75	2.57	[−47.97, 72.40]	.24
RP	5.10	1.76	2.91	[−17.20, 27.40]	.21
NS	2.99*	0.13	23.58	[1.38, 4.60]	.03
EE	5.18	2.26	2.29	[−2.53, 12.90]	.12
CCSSM practice standard[c]					
MP 2, 4, 5 with 1, 6	3.91	1.30	3.02	[−2.53, 10.36]	.11
MP 2, 4, 5 with 1, 6, 7	3.24**	0.46	7.13	[1.92, 4.57]	.003
MP 2, 4, 5 with 1, 3, 6	2.92	1.28	2.29	[−2.38, 8.22]	.14
MP 2, 4, 5 with 1, 3, 6, 7	4.72*	1.02	4.64	[1.57, 7.87]	.02

Note. WP = word problem; CI = confidence interval; CEC = Council for Exceptional Children; CCSSM = Common Core State Standards for Mathematics; NF = number and operations—fractions; G = geometry; RP = ratios and proportional relationships; NS = number system; EE = expressions and equations; MP = mathematical practice; OA = operations and algebraic thinking.
[a]The possible total CEC QI score was between 0 and 22 and was centered at the grand mean. [b]OA was the reference group. [c]MP 2, 4, and 5 was the reference group.
*p < .05. **p < .01. ***p < .001.

in solving one-variable equations (6.EE.B.7) and multistep operations of decimals (7.EE.B.3), NF (*n* = 3) in solving WPs with comparing and ordering fractions (4.NF.A.1, 4.NF.A.2) and multiplication and division of fractions (4.NF.B.4, 5.NF.B.4), G (*n* = 2) in solving area and perimeter of squares and rectangles (6.G.A.1, 7.G.B.6), NS (*n* = 2) in solving positive and negative integers (6.NS.C.5), and RP (*n* = 2) in applying the unit rate for multiplication and division of whole numbers and fractions (6.RP.A.2) or solving multiplication and division of percentages (6.RP.A.3).

As shown in Table 1, the effects of WP-solving instruction for students with LDs were large in general for all six CCSSM content standards, with a weighted average effect size estimate between 2.98 and 11.80. Although statistically nonsignificant, the intervention effects on G (BC-SMD = 11.80, *p* > .05) and EE (BC-SMD = 6.21, *p* > .05) indicated notably greater improvements than those on the other domains. The improvements from the baseline to intervention were statistically significant only for NS (BC-SMD = 3.04, *p* < .05) and OA (BC-SMD = 2.98, *p* < .001). The between-study heterogeneity (I^2) fell within a wide range (min = 51.50%, max = 93.16%, median = 78.53%, SD = 17.21).

In addition, as shown in Table 2, the effect of WP-solving instruction for students with LDs on the OA domain was smaller than that on all other observed content domains; this domain showed the largest deviation compared with G (β = 12.21, *p* = .24), followed by the EE (β = 5.18, *p* = .12), NF (β = 5.14, *p* = .17), RP (β = 5.10, *p* = .21), and NS (β = 2.99, *p* = .03). The intervention effect on WP solving related to NS was only statistically higher than OA.

Extent of Meeting CCSSM Practice Standards and Moderating Effect (RQ4)

The third moderator was the CCSSM practice standard (NGA Center & CCSSO, 2010). The results for the CCSSM practice standards are provided in Table S4. The average number of CCSSM practice standards used across the 20 studies was approximately six (min = 3, max = 7, median = 6, SD = 1.16). All studies focused on students' modeling with mathematics via pictorial, manipulative, verbal, and numeric representations (MP 4), quantitative reasoning by asking questions such as "What is the fraction used to describe the bar?" to explain their visual representations and clarify their thinking (MP 2), and using appropriate tools strategically, such as pencil and paper, iPad, and worksheets (MP 5). The next most frequently observed CCSSM practice standards included making sense of problems and checking themselves by asking questions such as "Does this make sense?" (MP 1, *n* = 18) and discussing with others with clear mathematical symbols and definitions with questions such as "What does the term 'numerator' mean?" (MP 6, *n* = 18). On the contrary, the use of constructing viable arguments with others, discussing why and how to use WP-solving steps (MP 3, *n* = 9), and discerning the patterns and structures of schematic diagrams for particular WP types (MP 7, *n* = 10) were relatively limited. Noticeably, no study has incorporated how to look for regularity in the repeated reasoning of WP situations.

As shown in Table 1, each weighted average effect size estimate (BC-SDM) of WP instruction for students with LDs was large across the seven CCSSM practice standards

of MP 1 through MP 7 (no study incorporated MP 8), ranging from 3.98 to 6.21. Except for MP 7, which showed the highest intervention effect (BC-SMD = 6.21, $p > .05$), the improvements from baseline to intervention phases were significant on MP 1 through MP 6 ($ps < .05$). The between-study heterogeneity (I^2) fell within a wide range (min = 74.70%, max = 93.16%, median = 81.72%, SD = 5.78).

Moreover, Table 2 shows the combinations of CCSSM practice standards that were present across studies. The implementation of MP 2, 4, 5 only had relatively smaller impacts on students' WP-solving performances than those of incorporating more CCSSM practice standards, MP 2, 4, 5 with 1, 6 (β = 3.91, p = .11); MP 2, 4, 5 with 1, 6, 7 (β = 3.24, p = .003); MP 2, 4, 5 with 1, 3, 6 (β = 2.92, p = .14); and MP 2, 4, 5 with 1, 3, 6, 7 (β = 4.72, p = .02). In particular, the combinations of MP 2, 4, 5 with 1, 6, 7 and MP 2, 4, 5 with 1, 3, 6, 7 resulted in statistically higher improvements than in the case of the use of MP 2, 4, and 5 alone.

Discussion

The purpose of this meta-analysis of WP instruction studies using SCD was to update and extend prior syntheses and meta-analyses of WP instruction for students with LDs. We conducted this meta-analysis to evaluate the overall effects of WP instruction for students with LDs. Another purpose was to examine the moderating effects of study quality addressed through CEC QIs and CCSSM content and practice standards (NGA Center & CCSSO, 2010).

Overall Effect

The findings indicate strong overall effects of WP instruction for students with LDs who experience difficulty solving WPs. Specifically, in all 20 studies, students with LDs displayed significant improvement from baseline to intervention phases, yielding a significantly large overall weighted effect size estimate (BC-SMD = 4.52). This result is consistent with that of previous meta-analyses (Jitendra et al., 2018; Peltier et al., 2018; Zheng et al., 2013). Furthermore, the large degree of between-study heterogeneity (I^2 = 81.72%) indicated the existence of variance across studies in effect sizes (Higgins et al., 2003). This finding also supported the need for evaluating associations among moderators (study quality and CCSSM standards) and variation among effect sizes.

Evaluation of Study Quality

We evaluated the extent and effects of the study quality by implementing the CEC (2014) QIs. In the field of special education, the reporting of QIs is important to determine which interventions constituted reliable research leading to positive impacts on student performance (B. G. Cook & S. C. Cook, 2013). Thus, when aggregating intervention outcomes

through a systematic review and a meta-analytic procedure, examining and reporting the study quality is an essential element. In general, the weighted average effect size estimate of WP instruction for students with LDs was large and significant for all eight CEC QI categories. More importantly, meeting all the three fidelity QIs (5.1. adherence, 5.2. dosage, and 5.3. regularity) resulted in the greatest effect (BC-SMD = 5.54). This finding corroborates the significance of implementation fidelity in SCD studies where researchers should systematically manipulate independent variables over time across participants (Horner et al., 2005). Although all 20 studies met implementation fidelity criteria related to dosage (such as instructional duration, frequency, and coverage), a relatively limited number of studies met the study quality of implementation fidelity related to its adherence, evaluating the critical elements of interventions (n = 16), and regularity, observing interventions across phases, interventionists, settings, or participants (n = 10). This failure to adequately collect and describe implementation fidelity in this study is consistent with previous studies that examined study quality using CEC QIs (S. C. Cook et al., 2020).

Evaluation of the CCSSM Content Standard

Solving WPs is one of the major hurdles for students with LDs across grade levels, and higher order thinking concepts and skills are even more challenging for secondary students with LDs (Cortiella & Horowitz, 2014). Regarding the CCSSM content standards (NGA Center & CCSSO, 2010), the average weighted effect sizes of WP instruction for students with LDs were large across different CCSSM content standards and, in particular, were greater in the domains of G (BC-SMD = 11.80) and EE (BC-SMD = 6.21) than in OA (BC-SMD = 2.98). Although insignificant, the findings indicate that the use of visual diagrams within the WP-solving context incorporated with a technology tool (i.e., video modeling) was largely effective in supporting students' WPs with area and perimeter of squares and rectangles (Satsangi, Hammer, & Bouck, 2020; Satsangi, Hammer, & Hogan, 2019). Despite this progress, a limited number of studies focused on the advanced level of WPs within the domains of G (n = 2), NS (n = 2), and RP (n = 2). Considering that many secondary students with LDs consistently experience difficulty with topics such as ratios, proportions, and percentages at the secondary school level (Soares et al., 2018), more studies should fill this gap in our knowledge in this field.

Evaluation of the CCSSM Practice Standard

The emphasis on MP standards is not new. Since the release of *An Agenda for Action* in 1980 by the National Council of Teachers of Mathematics, mathematical problem-solving reasoning has become an essential element of school mathematics across K–12 (Koestler et al., 2013). Regarding the

latest MP standards published by the NGA Center and CCSSO (2010), the CCSSM initiative is aligned with the emphasis on students' quantitative reasoning and problem-solving processes. The current meta-analysis also supports that WP-solving instruction for students with LDs was largely effective across CCSSM practice standards in general (BC-SMD = 3.98–6.21). The effect sizes were particularly large to *look for and make use of the structure* (MP 7, BC-SMD = 6.21). This finding is consistent with that of several previous studies that demonstrated strong effects of SBI on teaching WPs to students with disabilities (Jitendra et al., 2018; Peltier et al., 2018). Moreover, all the reviewed studies incorporated *reasoning abstractly and quantitatively* (MP 2), *modeling with mathematics* (MP 4), and *using appropriate tools strategically* (MP 5) in their WP instruction and, to some degree, included additional components of CCSSM practice standards. In this study, the greater the number of MPs (greater than three) provided, the larger the effect size. This finding is consistent with a previous meta-analysis by Codding et al. (2011); however, Kim et al. (2020) found no distinctive relationship between these two factors. Therefore, we should continue to study how to effectively deliver each component with fidelity rather than relying on the use of multiple components.

Limitations and Future Research

In this meta-analysis, we systematically identified peer-reviewed SCD studies published between 1975 and 2020 for students with LDs in Grades 1 through 12. Despite the rigorous steps applied to the analysis of these studies over four decades, we noted four limitations when interpreting findings.

First, although we focused on students identified as having LDs, it is important to acknowledge there could be variance among samples. Although all 82 participants showed difficulty in WPs based on their screening tests, authors reported a subset of 16 students to have additional reading or writing difficulties. Students' types and levels of difficulty may vary within the group. Furthermore, students with LDs might have varying performances based on how they were identified as having LDs over the last four decades. We noticed that in studies published in before 2003, authors identified students as having LDs based on IQ-achievement discrepancy and/or performing more than 2 years below their grade levels. In studies published after 2002, authors identified students as having LDs based on various criteria such as IQ-achievement discrepancy, performing below the 10th or 25th percentile on a standardized mathematics test, or school's response to intervention (RTI) process. Thus, these factors could have influenced variances within samples and require caution in interpreting the findings.

Second, although CEC (2014) QIs are frequently used to evaluate study quality in the field of special education (S. C. Cook et al., 2020), the use of different QI guidelines (WWC,

2020) can yield different results. Furthermore, we have only included SCD studies with at least three participants with a minimum of three data points per phase, excluding studies with AB[k] or alternating treatment designs without baseline or multiple-baseline/multiple-probe across behaviors or settings. These procedures could inflate the publication bias in the present meta-analysis by excluding data that did not meet the rigorous inclusion criteria. Researchers should continuously validate study quality across different QI guidelines and develop data analytic tools across various types of SCD to remedy these methodological limitations.

Third, we located a limited number ($n = 20$) of studies to examine the effects of WPs on the mathematics performance of students with LDs. Due to this relatively small number of studies, the results of some moderator analyses were based on a small number of effect sizes. The insufficient number of studies might have underpowered the moderating effects in meta-regression (Borenstein et al., 2009). Furthermore, no researchers in the included studies incorporated the CCSSM practice standard of looking for and expressing regularity in repeated reasoning (MP 8). To help students generalize mathematical patterns and arguments through reasoning and the proof process, teachers can ask such questions as "How did you determine the answer?" or "Do you think that will always work?" and help students flexibly make connections between identified patterns and algorithms in WP solving (Koestler et al., 2013). Thus, future studies on the moderators that were not investigated in this study, such as the CCSSM practice standard MP 8, are needed to further validate the moderating effects.

Finally, because the SCD data are collected over time under similar conditions, the assumption of independence is easily violated (Wolery et al., 2010). Thus, in the current meta-analysis, we applied a multivariate analytic method accounting for dependence between outcomes over time, considering between-case variances (Moeyaert et al., 2020; Pustejovsky et al., 2014). In future SCD meta-analysis studies, researchers should continuously validate the synthesizing methods for SCD data and easily interpret the results for educators and practitioners in the field.

Implications for Practice

The findings in the current meta-analysis indicated that the study quality (CEC, 2014) and the combination of CCSSM standards (NGA Center & CCSSO, 2010) are essential considerations in designing WP interventions for students with LDs. WP instruction showed a large and significant overall effect for students with LDs. We encourage teachers to employ WP instruction when teaching various WP topics to students with LDs who experience difficulty representing and solving WPs (Bryant et al., 2000). Given the importance of being able to solve WPs successfully and proficiently, students with LDs must benefit from WP instruction to enable them to keep pace with their same-grade peers.

Furthermore, the magnitude of the average weighted effect sizes of WP instruction was likely to have been influenced by how well each study met the quality criteria, focused on different CCSSM content standards, and incorporated different CCSSM practice standards in mathematics interventions. Meeting the CEC QI of the implementation of fidelity showed a substantial increase in the overall effect. Thus, when teachers provide WP instruction to students, it is important to consistently maintain rigor in the scope and quality of interventions in their day-to-day practice. The current findings also indicate relatively higher effects of WP instruction for the secondary grade-level (e.g., G, EE, and RP) than the elementary grade-level (e.g., OA) mathematics domains.

Finally, the findings of this meta-analysis indicate that using multiple MPs does not guarantee higher intervention effects. The number and combination patterns of CCSSM practice standards employed in the interventions varied, resulting in nonlinear relationships with effect sizes; thus, we cannot guarantee that higher numbers of instructional components lead to higher intervention effects. Ultimately, when incorporating MPs in their lessons, teachers should appropriately utilize each instructional component for teaching WPs to students, particularly those with LDs. Teachers should consider mathematical domain-specific visual models and check whether students have sufficient opportunities for review and practice.

Declaration of Conflicting Interests

Funding

ORCID iD

Mikyung Shin https://orcid.org/0000-0001-7907-9193

Supplemental Material

Supplemental material for this article is available online.

References

Altman, D. G. (1999). *Practical statistics for medical research*. Chapman & Hall/CRC Press.

Andersson, U. (2008). Mathematical competencies in children with different types of learning difficulties. *Journal of Educational Psychology*, *100*(1), 48–66. https://doi.org/10.1037/0022-0663.100.1.48

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3–25. https://doi.org/10.1037/amp0000191

Araujo, J., & Born, D. G. (1985). Calculating percentage agreement correctly but writing its formula incorrectly. *The Behavior Analyst*, *8*(2), 207–208. https://doi.org/10.1007/bf03393152

Baek, E. K., Petit-Bois, M., Van den Noortgate, W., Beretvas, S. N., & Ferron, J. M. (2016). Using visual analysis to evaluate and refine multilevel models of single-case studies. *The Journal of Special Education*, *50*(1), 18–26. https://doi.org/10.1177/0022466914565367

Boonen, A. J., de Koning, B. B., Jolles, J., & van der Schoot, M. (2016). Word problem solving in contemporary math education: A plea for reading comprehension skills training. *Frontiers in Psychology*, *7*, Article 191. https://doi.org/10.3389/fpsyg.2016.00191

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Meta-regression. In M. Borenstein, L. V. Hedges, J. P. T. Higgins, & H. R. Rothstein (Eds.), *Introduction to meta-analysis* (pp. 187–203). Wiley.

Bryant, D. P., Bryant, B. R., & Hammill, D. D. (2000). Characteristic behaviors of students with learning disabilities who have teacher-identified math weaknesses. *Journal of Learning Disabilities*, *33*(2), 168–177. https://doi.org/10.1177/002221940003300205

Codding, R. S., Burns, M. K., & Lukito, G. (2011). Meta-analysis of mathematic basic-fact fluency interventions: A component analysis. *Learning Disabilities Research & Practice*, *26*(1), 36–47. https://doi.org/10.1111/j.1540-5826.2010.00323.x

Cook, B. G., & Cook, S. C. (2013). Unraveling evidence-based practices in special education. *The Journal of Special Education*, *47*(2), 71–82. https://doi.org/10.1177/0022466911420877

Cook, S. C., Collins, L. W., Morin, L. L., & Riccomini, P. J. (2020). Schema-based instruction for mathematical word problem solving: An evidence-based review for students with learning disabilities. *Learning Disability Quarterly*, *43*(2), 75–87. https://doi.org/10.1177/0731948718823080

Cortiella, C., & Horowitz, S. H. (2014). *The state of learning disabilities: Facts, trends and emerging issues*. National Center for Learning Disabilities.

Council for Exceptional Children. (2014). *CEC standards for evidence-based practices in special education*.

Dennis, M. S., Knight, J., & Jerman, O. (2016). Teaching high school students with learning disabilities to use model drawing strategy to solve fraction and percentage word problems. *Preventing School Failure: Alternative Education for Children and Youth*, *60*(1), 10–21. https://doi.org/10.1080/1045988X.2014.954514

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Fisher, Z., Tipton, E., & Zhipeng, H. (2017). *robumeta: Robust variance meta-regression* (R package version 2.0) [Computer software]. https://github.com/zackfisher/robumeta

Fuchs, L. S., & Fuchs, D. (2002). Mathematical problem-solving profiles of students with mathematics disabilities with and without comorbid reading disabilities. *Journal of Learning Disabilities*, *35*(6), 564–574. https://doi.org/10.1177/00222194020350060701

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with

learning disabilities: A meta-analysis of instructional components. *Review of Educational Research*, *79*(3), 1202–1242. https://doi.org/10.3102/0034654309334431

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, *71*(2), 149–164. https://doi.org/10.1177/001440290507100202

GetData Graph Digitizer. (2013). *GetData graph digitizer* (Version 2.26) [Computer software]. http://getdata-graph-digitizer.com

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. https://doi.org/10.1002/jrsm.5

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, *71*(2), 165–179. https://doi.org/10.1177/001440290507100203

Hwang, J., & Riccomini, P. J. (2016). Enhancing mathematical problem solving for secondary students with or at risk of learning disabilities: A literature review. *Learning Disabilities Research & Practice*, *31*(3), 169–181. https://doi.org/10.1111/ldrp.12105

Jitendra, A. K., Lein, A. E., Im, S.-H., Alghamdi, A. A., Hefte, S. B., & Mouanoutoua, J. (2018). Mathematical interventions for secondary students with learning disabilities and mathematics difficulties: A meta-analysis. *Exceptional Children*, *84*(2), 177–196. https://doi.org/10.1177/0014402917737467

Jitendra, A. K., Nelson, G., Pulles, S. M., Kiss, A. J., & Houseworth, J. (2016). Is mathematical representation of problems an evidence-based strategy for students with mathematics difficulties? *Exceptional Children*, *83*(1), 8–25. https://doi.org/10.1177/0014402915625062

Jitendra, A. K., Petersen-Brown, S., Lein, A. E., Zaslofsky, A. F., Kunkel, A. K., Jung, P., & Egan, A. M. (2015). Teaching mathematical word problem solving: The quality of evidence for strategy instruction priming the problem structure. *Journal of Learning Disabilities*, *48*(1), 51–72. https://doi.org/10.1177/0022219413487408

Jitendra, A. K., & Xin, Y. P. (1997). Mathematical word-problem-solving instruction for students with mild disabilities and students at risk for math failure: A research synthesis. *The Journal of Special Education*, *30*(4), 412–438. https://doi.org/10.1177/002246699703000404

Jordan, N. C., & Hanich, L. B. (2000). Mathematical thinking in second-grade children with different forms of LD. *Journal of Learning Disabilities*, *33*(6), 567–578. https://doi.org/10.1177/002221940003300605

Keith, T. Z. (2019). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling* (3rd ed.). Routledge.

Kim, S. A., Bryant, P. D., Bryant, R. B., Shin, M., Ok, M. W., & Nozari, M. (2020). *A synthesis of number combinations and whole number computation interventions for students with mathematics disabilities*. Manuscript submitted for publication.

Kingsdorf, S., & Krawec, J. (2014). Error analysis of mathematical word problem solving across students with and without learning disabilities. *Learning Disabilities Research & Practice*, *29*(2), 66–74. https://doi.org/10.1111/ldrp.12029

Koestler, C., Felton-Koestler, M. D., Bieda, K., & Otten, S. (2013). *Connecting the NCTM process standards and the CCSSM practices*. National Council of Teachers of Mathematics.

Lein, A. E., Jitendra, A. K., & Harwell, M. R. (2020). Effectiveness of mathematical word problem solving interventions for students with learning disabilities and/or mathematics difficulties: A meta-analysis. *Journal of Educational Psychology*, *112*(7), 1388–1408. https://doi.org/10.1037/edu0000453

Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science*, *26*, 49–63. https://doi.org/10.1023/A:1003088013286

Moeyaert, M., Manolov, R., & Rodabaugh, E. (2020). Meta-analysis of single-case research via multilevel models: Fundamental concepts and methodological considerations. *Behavior Modification*, *44*(2), 265–295. https://doi.org/10.1177/0145445518806867

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, *151*(4), 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Montague, M., & Dietz, S. (2009). Evaluating the evidence base for cognitive strategy instruction and mathematical problem solving. *Exceptional Children*, *75*(3), 285–302. https://doi.org/10.1177/001440290907500302

Montague, M., Enders, C., & Dietz, S. (2011). Effects of cognitive strategy instruction on math problem solving of middle school students with learning disabilities. *Learning Disability Quarterly*, *34*(4), 262–272. https://doi.org/10.1177/0731948711421762

National Council of Teachers of Mathematics. (2014). *Principles to actions: Ensuring mathematical success for all*.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*.

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

Odom, S. L., Barton, E. E., Reichow, B., Swaminathan, H., & Pustejovsky, J. E. (2018). Between-case standardized effect size analysis of single case designs: Examination of the two methods. *Research in Developmental Disabilities*, *79*, 88–96. https://doi.org/10.1016/j.ridd.2018.05.009

Peake, C., Jiménez, J. E., Rodríguez, C., Bisschop, E., & Villarroel, R. (2015). Syntactic awareness and arithmetic word problem solving in children with and without learning disabilities. *Journal of Learning Disabilities*, *48*(6), 593–601. https://doi.org/10.1177/0022219413520183

Peltier, C. J., Vannest, K. J., & Marbach, J. J. (2018). A meta-analysis of schema instruction implemented in single-case experimental designs. *The Journal of Special Education*, *52*(2), 89–100. https://doi.org/10.1177/0022466918763173

Pfannenstiel, K. H., Bryant, D. P., Bryant, B. R., Porterfield, J. A. (2015). Cognitive strategy instruction for teaching word problems to primary-level struggling students. *Intervention in School and Clinic*, *50*(5), 291–296. https://doi.org/10.1177/1053451214560890

Pinheiro, J. C., Bates, D. M., DebRoy, S., & Sarkar, D. (2020). *nlme: Linear and nonlinear mixed effects models* (R package version 3.1-145) [Computer software]. https://svn.r-project.org/R-packages/trunk/nlme

Powell, S. (2011). Solving word problems using schemas: A review of the literature. *Learning Disabilities Research & Practice*, *26*(2), 94–108. https://doi.org/10.1111/j.1540-5826.2011.00329.x

Pustejovsky, J. E. (2016). *scdhlm: Estimating hierarchical linear models for single-case designs* (R package version 0.3.1) [Computer software]. https://github.com/jepusto/scdhlm

Pustejovsky, J. E., Hedges, L. V., & Shadish, W. R. (2014). Design-comparable effect sizes in multiple baseline designs: A general modeling framework. *Journal of Educational and Behavioral Statistics*, *39*(5), 368–393. https://doi.org/10.3102/1076998614547577

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Wiley.

Satsangi, R., Hammer, R., & Bouck, E. C. (2020). Using video modeling to teach geometry word problems: A strategy for students with learning disabilities. *Remedial and Special Education*, *41*(5), 309–320. https://doi.org/10.1177/0741932518824974

Satsangi, R., Hammer, R., & Hogan, C. D. (2019). Video modeling and explicit instruction: A comparison of strategies for teaching mathematics to students with learning disabilities. *Learning Disabilities Research & Practice*, *34*(1), 35–46. https://doi.org/10.1111/ldrp.12189

Shadish, W. R., Hedges, L. V., Horner, R. H., & Odom, S. L. (2015). *The role of between-case effect size in conducting, interpreting, and summarizing single-case research* (NCER 2015-002). National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

Shin, M., & Bryant, D. P. (2015a). Fraction interventions for students struggling to learn mathematics: A research synthesis. *Remedial and Special Education*, *36*(6), 374–387. https://doi.org/10.1177/0741932515572910

Shin, M., & Bryant, D. P. (2015b). A synthesis of mathematical and cognitive performances of students with mathematics learning disabilities. *Journal of Learning Disabilities*, *48*(1), 96–112. https://doi.org/10.1177/0022219413508324

Shin, M., Bryant, D. P., Powell, S. R., Jung, P.-G., Ok, M. W., & Hou, F. (2020). *A meta-analysis of single-case research on word-problem instruction for students with learning disabilities* [Data files and scripts]. https://doi.org/10.17605/OSF.IO/XR58S

Shin, M., & Jung, P. (2018). Reliability, validity, and usability of a graph extraction computer program for the meta-analysis of single-case studies. *Special Education Research*, *17*(3), 185–20. https://doi.org/10.18541/ser.2018.08.17.3.185

Soares, N., Evans, T., & Patel, D. R. (2018). Specific learning disability in mathematics: A comprehensive review. *Translational Pediatrics*, *7*(1), 48–62. https://doi.org/10.21037/tp.2017.08.03

Spooner, F., Saunders, A., Root, J., & Brosh, C. (2017). Promoting access to Common Core Mathematics for students with severe disabilities through mathematical problem solving. *Research and Practice for Persons with Severe Disabilities*, *42*(3), 171–186. https://doi.org/10.1177/1540796917697119

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*(3), 375–393. https://doi.org/10.1037/met0000011

Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, *40*(6), 604–634. https://doi.org/10.3102/1076998615606099

Tolar, T. D., Fuchs, L., Fletcher, J. M., Fuchs, D., & Hamlett, C. L. (2016). Cognitive profiles of mathematical problem solving learning disability for different definitions of disability. *Journal of Learning Disabilities*, *49*(3), 240–256. https://doi.org/10.1177/0022219414538520

van Garderen, D., Scheuermann, A., & Poch, A. (2014). Challenges students identified with a learning disability and as high-achieving experience when using diagrams as a visualization tool to solve mathematics word problems. *ZDM Mathematics Education*, *46*, 135–149. https://doi.org/10.1007/s11858-013-0519-1

Viswanathan, M., Patnode, C. D., Berkman, N. D., Bass, E. B., Chang, S., Hartling, L., Murad, M. H., Treadwell, J. R., & Kane, R. L. (2018). Recommendations for assessing the risk of bias in systematic reviews of health-care interventions. *Journal of Clinical Epidemiology*, *97*, 26–34. https://doi.org/10.1016/j.jclinepi.2017.12.004

Vukovic, R. K., Lesaux, N. K., & Siegel, L. S. (2010). The mathematics skills of children with reading difficulties. *Learning and Individual Differences*, *20*(6), 639–643. https://doi.org/10.1016/j.lindif.2010.08.004

Wei, X., Lenz, K. B., & Blackorby, J. (2013). Math growth trajectories of students with disabilities: Disability category, gender, racial, and socioeconomic status differences from ages 7 to 17. *Remedial and Special Education*, *34*(3), 154–165. https://doi.org/10.1177/0741932512448253

What Works Clearinghouse. (2011). *WWC procedures and version 2.1 standards handbook*. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v2_1_standards_handbook.pdf

What Works Clearinghouse. (2020). *WWC procedures and version 4.1 standards handbook*. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf

Wolery, M., Busick, M., Reichow, B., & Barton, E. E. (2010). Comparison of overlap methods for quantitatively synthesizing single-subject data. *The Journal of Special Education*, *44*(1), 18–28. https://doi.org/10.1177/0022466908328009

Xin, Y. P., & Jitendra, A. K. (1999). The effects of instruction in solving mathematical word problems for students with learning problems: A meta-analysis. *Journal of Special Education*, *32*(4), 207–225. https://doi.org/10.1177/002246699903200402

Zheng, X., Flynn, L. J., & Swanson, H. L. (2013). Experimental intervention studies on word problem solving and math disabilities: A selective analysis of the literature. *Learning Disability Quarterly*, *36*(2), 97–111. https://doi.org/10.1177/0731948712444277