

Three-Level Meta-Analysis of Single-Case Data Regarding the Effects of Peer Tutoring on Academic and Social-Behavioral Outcomes for At-Risk Students and Students With Disabilities

Remedial and Special Education
2021, Vol. 42(2) 94–106
© Hammill Institute on Disabilities 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0741932519855079
rase.sagepub.com



Mariola Moeyaert, PhD¹, David A. Klingbeil, PhD² ,
Emily Rodabaugh, MSc¹, and Merve Turan, MSc¹ 

Abstract

Meta-analysis of single-case experimental designs may further knowledge about evidence-based practices for students needing remedial or special education. To contribute to evidence-based practice, a multivariate multilevel meta-analysis was used to synthesize the effectiveness of peer tutoring interventions on both academic and social-behavior outcomes. In total, 46 single-case studies met all inclusion criteria. Peer tutoring had a statistically significant effect on both academic and social-behavior outcomes, with a slightly larger effect on academic outcomes. Peer tutoring also had a significant effect on the trend in academic outcomes during the treatment phase (indicating that the intervention becomes more effective over time), but the effect on trends was slightly less than for social outcomes. Including moderators such as gender, age, disability type, and study quality reduced the amount of between-case and between-study heterogeneity. Limitations and implications of these findings are discussed.

Keywords

multivariate multilevel modeling, single-case experimental design, peer tutoring

The use of peer-mediated interventions has a long-standing history in educational settings (Cohen, Kulik, & Kulik, 1982; Devin-Sheehan, Feldman, & Allen, 1976). Peer-mediated interventions include peer modeling, peer management strategies (which includes peer initiation training, peer monitoring, and peer coaching), peer network training, group-oriented contingencies, and peer tutoring (Kohler & Strain, 1990; Utley & Mortweet, 1997). Researchers have typically defined peer tutoring as an intervention in which peers provide one-on-one instruction, practice opportunities, and feedback to a target student on academic skill responses. There is a large body of evidence supporting the use of peer tutors to improve the academic achievement of at-risk students and students with disabilities (e.g., Cohen et al., 1982; Cook, Scruggs, Mastropieri, & Casto, 1985; Mastropieri, Spencer, Scruggs, & Talbott, 2000; Mathes & Fuchs, 1994; Okilwa & Shelby, 2010). However, most meta-analytic summaries of peer tutoring have excluded single-case designs that are commonly used in special education (Horner et al., 2005). The purpose of this study was to conduct a comprehensive, methodologically rigorous meta-analytic review of peer tutoring used with at-risk students and for students with disabilities.

Peer tutoring incorporates several principles of effective instruction including scaffolding, increased opportunities to respond, repetition, and immediate corrective feedback (Utley & Mortweet, 1997). By increasing students' access to these empirically supported instructional principles, peer tutoring is a systematic method to provide additional support and increase students' academic skills. Peer tutoring is also an efficient method for providing additional supports to students within a tiered service delivery framework.

There is substantial evidence that academic and behavior problems often co-occur (Darney, Reinke, Herman, Stormont, & Jalongo, 2013; Hinshaw, 1992; King, Gonzales, & Reinke, 2018). Although other peer-mediated interventions have been used to directly address the target student's

¹University at Albany–State University of New York, USA

²University of Texas at Austin, USA

Corresponding Author:

Mariola Moeyaert, Department of Educational and Counseling Psychology, University at Albany–State University of New York, 1400 Washington Ave., Albany, NY 12222, USA.
Email: mmoeyaert@albany.edu

Associate Editor: Daniel Maggin

behavior (e.g., Dart, Collins, Klingbeil, & McKinley, 2014), there is theoretical and empirical support for the indirect effects of academic peer tutoring on social-behavioral outcomes. For example, escape from difficult work is a common motivating consequence for problem behavior (O'Connor & Daly, 2018). Therefore, improving students' academic skills may impact the antecedents that control the problem behavior. Tutoring may also increase access to prosocial models and serve to prompt verbal responses. Indeed, researchers have found that peer tutoring targeting academic skills has led to decreases in off-task and noncompliant behavior (DuPaul, Ervin, Hook, & McGoey, 1998; McDonnell, Mathot-Buckner, Thorson, & Fister, 2001) as well as increased academic engagement (Ginsburg-Block & Fantuzzo, 1997), positive social interactions (e.g., Maheady & Sainato, 1985), and the social acceptance of targeted students (Fuchs, Fuchs, Mathes, & Martinez, 2002).

Peer tutoring interventions vary on several aspects. Examples of peer tutoring at the universal level include classwide peer tutoring (Delquadri, Greenwood, Whorton, Carta, & Hall, 1986; Greenwood, Delquadri, & Hall, 1989) and Peer Assisted-Learning Strategies (PALS; Fuchs, Fuchs, Mathes, & Simmons, 1997). Both of these programs incorporate reciprocal peer tutoring during which two students are paired and each student takes a turn serving as the tutor. Notably, the PALS program requires the pairing of high- and low-achieving students. Reciprocal tutoring procedures differ from nonreciprocal peer tutoring wherein the tutor and target student do not change roles (Mastropieri et al., 2000). Researchers have investigated the use of nonreciprocal peer tutors who were older than the target students and peer tutors of the same age (Cohen et al., 1982). Moreover, research has found that students with disabilities can serve as effective tutors (e.g., Cook et al., 1985). Readers interested in a more thorough review of the types of peer tutoring are referred to Talbott, Trzaska, and Zurheide (2017).

Previous Reviews of Peer Tutoring

A number of systematic reviews and meta-analytic syntheses regarding peer tutoring exist, with the majority of meta-analytic studies focusing on group designs. There is evidence to support the use of peer tutoring broadly (Cohen et al., 1982) and for specific tutoring programs such as PALS (McMaster, Fuchs, & Fuchs, 2006; Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003). There is also evidence to support the use of peer tutoring for students (a) identified as at risk for academic difficulties (Wexler, Reed, Pyle, Mitchell, & Barton, 2015); (b) mild (or high incidence) disabilities including behavior disorders, learning disabilities, or mild intellectual disability (Cook et al., 1985; Mathes & Fuchs, 1994); and (c) low incidence disabilities including autism (Bene, Banda, & Brown, 2014) and

intellectual disability (Schaefer, Cannella-Malone, & Carter, 2016). However, the effects of peer tutoring programs are not uniform. For example, McMaster et al. (2006) systematically reviewed the PALS literature and found limited effects for students with disabilities, particularly students' with deficits in attention, behavioral control, and cognitive development (e.g., Al Otaiba & Fuchs, 2002).

The quantitative synthesis of single-case research on peer tutoring is less common. Single-case experimental design (SCED) studies provide a methodologically rigorous method for documenting intervention effectiveness that is well suited for educational and clinical settings (Kazdin, 2011; Kratochwill et al., 2010). Excluding single-case research from research syntheses represents an important gap in the literature due to the prevalence of SCEDs conducted with individuals with disabilities (Horner et al., 2005). Meta-analytic reviews of single-case research may provide additional information regarding the generalizability of the findings (Maggin, 2015), perhaps especially for at-risk youth and youth with disabilities. Recent methodological advances in quantifying effect sizes from SCEDs and synthesizing effect sizes across SCED studies could begin to fill this gap in the literature by identifying key variables that moderate the effectiveness of peer tutoring for youth with disabilities. When synthesizing effect sizes (reflecting the effectiveness of peer tutoring in this study) across a variety of studies measuring the same outcome variable, there is more evidence regarding the effectiveness of the intervention.

In an early study that included single-case effect sizes, Stenhoff and Lignugaris/Kraft (2007) reviewed 12 SCED studies of peer tutoring for secondary students with mild disabilities. Using the percentage of nonoverlapping data (PND) to quantify treatment effects, the authors found evidence that peer tutoring led to improvements in correct academic responding ($M = 74\%$). Not enough SCED studies included measures of social-behavioral outcomes such as on-task behavior and communication to calculate average effect sizes. Ryan, Reid, and Epstein (2004) reviewed peer tutoring studies within a larger review of peer-mediated interventions for students with emotional or behavioral disorders. Ryan et al. (2004) mixed results from 14 group design and SCED studies and reported smaller effect sizes for reading (0.82) compared with math (2.08), English (2.03), or history (3.00). Larger effects were found for same-age peer tutors (1.92) than cross-age peer tutors (1.12).

Bowman-Perrott et al. (2013) found that peer tutoring had medium effects on academic outcomes ($\text{Tau-U} = .76$) in 26 peer-reviewed published between 1966 and 2011. Peer tutoring had larger effects for vocabulary ($\text{Tau-U} = .92$) and math ($\text{Tau-U} = .86$) than for reading ($\text{Tau-U} = .77$), spelling ($\text{Tau-U} = .74$), or social studies ($\text{Tau-U} = 0.57$). The effects of peer tutoring were not significantly

moderated by grade level (e.g., elementary vs. secondary) or dosage (split at the median number of treatment minutes). Finally, larger effects were found for students who were at risk or had disabilities (Tau-U = .76) than non-at risk or nondisabled students (Tau-U = 0.65). Similar effect sizes were reported for students with learning disabilities (Tau-U = 0.75) and emotional and behavioral disorder (Tau-U = 0.76).

Finally, Bowman-Perrott and colleagues synthesized the effects of “peer tutoring as an intervention (with or without an academic component) used to address social or behavioral outcomes” (Bowman-Perrott, Burke, Zhang, & Zaini, 2014). In 20 studies that met WWC criteria (Kratochwill et al., 2010) published between 1966 and 2012, peer tutoring interventions had a medium overall effect (Tau-U = 0.62). Similar effects were found for increasing target students’ social interactions and social skills (Tau-U = 0.69) and decreasing their disruptive or off-task behaviors (Tau-U = 0.60), whereas smaller effects were found for academic engagement (Tau-U = 0.38). Peer tutoring effects varied based on student disability status. Effects were largest for students with intellectual disability ($n = 5$; Tau-U = .93) followed by students with other health impairments, including attention-deficit disorder or attention-deficit/hyperactivity disorder (ADHD); $n = 23$; Tau-U = .63, emotional and behavioral disorders ($n = 55$; Tau-U = .61), learning disabilities ($n = 24$; Tau-U = .57), and autism ($n = 5$; Tau-U = .49), respectively. Finally, peer tutoring interventions which targeted social-behavioral outcomes directly ($k = 14$) were associated with larger effects (Tau-U = 0.75) compared with peer tutoring targeting academic effects primarily ($k = 6$; Tau-U = 0.43). There was a positive correlation between academic outcome gains and increased academic engagement ($r = 0.52$) and reductions in disruptive or off-task behavior ($r = 0.31$). However, there was a negative correlation between academic outcome gains and increased social interactions and social skills ($r = -.18$; Bowman-Perrott et al., 2014).

The meta-analytic reviews which included SCEDs are laudable but not without limitations. The use of more recently developed techniques for synthesizing single-case research may provide a more nuanced synthesis of the effectiveness of peer tutoring for youth with disabilities. First, there were issues with the effect sizes used. The first two studies used effect sizes (Ryan et al., 2004; Stenhoff & Lignugaris/Kraft, 2007) with undesirable or unknown statistical properties (Shadish, 2014). Bowman-Perrott et al. (2014; Bowman-Perrott et al., 2013) estimated effect sizes by calculating Tau-U (Parker, Vannest, Davis, & Sauber, 2011) which improved upon earlier nonoverlap effect sizes but is still limited. Tau-U is purported to represent the percentage of non-overlapping data minus the percentage of overlapping data, meaning that Tau-U values should be bound between -1 and 1 (Parker et al., 2011). However, when Tau-U is applied to

single-case data, the obtained values often exceed these bounds (Klingbeil, Van Norman, McLendon, Ross, & Begeny, 2019). According to Tarlow (2017), this issue is due to an alteration Parker et al. (2011) made to the original formula for estimating tau. As a result, Tau-U values are often positively biased (i.e., intervention effects are over estimated) and difficult to interpret (Tarlow, 2017).

Second, hierarchical linear modeling, as an extension of the regression-based effect sizes, captures the nesting structure inherent in single-case research designs. Three-level hierarchical linear models allow researchers to model variation between observations within an individual, variation between individuals within a study, and variation between studies (Moeyaert, Ugille, Ferron, Beretvas, & Van den Noortgate, 2013a, 2013b; Van den Noortgate & Onghena, 2008). A three-level model may be a more appropriate method of examining the effect of participant-level moderators (e.g., disability status) compared with aggregating at the study level. In addition, traditional meta-analyses do not differentiate between the case and the study level. As a consequence, if there is a lot of variance found in treatment effects, then only study-level moderators or aggregated-case moderators (such as the average age per study) can be added. Third, single-case researchers often measure intervention effects on multiple measures (e.g., academic and social-behavioral outcomes). To account for the dependency between outcomes, Bowman-Perrott et al. (2014) and Bowman-Perrott et al. (2013) calculated an average effect on each outcome for each study and afterward combined these study-level average using traditional meta-analytic procedures. Multivariate approaches, inherently modeling the dependency between outcome scores, provide more accurate effect size estimates (Hedges, Tipton, & Johnson, 2010). Therefore, a multivariate multilevel model is more appropriate.

Fourth, two studies focused on students with specific types of disabilities (Ryan et al., 2004; Stenhoff & Lignugaris/Kraft, 2007), whereas Bowman-Perrott et al. (2013) aggregated results for students who were at risk or disabled compared with their not at risk or nondisabled peers. This is potentially problematic as research suggests peer tutoring may have differential effects based on the severity of disability (McMaster et al., 2006). Bowman-Perrott et al. (2014) evaluated differential effects for six different disability categories. However, that meta-analysis also included studies of peer-mediated or peer-management interventions (Kohler & Strain, 1990). The authors did not report whether there were differential effects of peer tutoring based on the students’ disability categories.

Moderators of Peer Tutoring Effectiveness

In addition to the SCED research reviewed above, previous systematic reviews and meta-analytic of group design

research provide further insight into the types of variables that may moderate the effectiveness of peer tutoring. What follows is a brief review of previous meta-analytic research conducted with group designs. We included results from reviews that did not target students with disabilities to provide a more comprehensive picture of the variability in previous research.

Targeted Outcome

Multiple meta-analytic reviews have found that peer tutoring has positive effects on overall academic achievement (e.g., Cohen et al., 1982; Rohrbeck et al., 2003). However, the effect of peer tutoring may differ depending on the targeted outcome. Cook et al. (1985) found larger effects for peer tutoring interventions for children with disabilities focused on language and math when compared with interventions teaching reading. Similarly, Leung's (2015) comprehensive meta-analysis found that peer tutoring had larger effects on other subjects (e.g., physical education, arts, science) than either math or reading. However, when only synthesizing results from standardized tests, peer tutoring had larger effects on math and reading compared with the other subjects. This suggests that the type of test may have influenced the observed differences in peer tutoring effects between academic skill outcomes. Evidence also suggests that peer tutoring can have positive effects on behavioral outcomes. For example, Cook et al. (1985) found that peer tutoring had small indirect effects on the behavioral functioning of children with disabilities.

Participant Characteristics

Various participant characteristics may impact the effects peer tutoring. Some evidence suggests that peer tutoring may be more effective in elementary grades (Cohen et al., 1982; Rohrbeck et al., 2003) although this finding was not replicated in more recent work (Leung, 2015). Gender matching also appears to impact peer tutoring effectiveness. Peer tutoring has been associated with larger effect sizes in studies using gender-matched intervention dyads than mixed gender dyads (Leung, 2015; Rohrbeck et al., 2003). The same effect has not been found for age matching with evidence supporting same-age and cross-age peer tutors (Jun, Ramirez, & Cumming, 2010; Leung, 2015; Mathes & Fuchs, 1994). Tutor skill may also be an important moderator of peer tutoring effectiveness. Leung (2015) found that studies using peer tutors with low ability were associated with larger effect sizes than studies with tutors with high or average ability.

Evidence supports the use of peer tutoring with low-achieving youth (Elbaum, Vaughn, Tejero Hughes, & Watson Moody, 2000; Leung, 2015) and youth with disabilities (Cook et al., 1985). Published systematic reviews

support the use of peer tutoring with children with autism spectrum disorders (Bene et al., 2014), intellectual disability (Schaefer et al., 2016), learning disabilities (Kunsch, Jitendra, & Sood, 2007; Mathes & Fuchs, 1994), and behavior disorders (Ryan et al., 2004; Spencer, 2006). However, much less is known about the differential effects of peer tutoring for children with other disabilities.

Study Quality

Study quality may also impact the observed effects of peer tutoring. For example, Cook et al. (1985) found smaller overall effects in studies that had fewer threats to internal validity (i.e., higher quality studies). However, Ginsburg-Block, Rohrbeck, and Fantuzzo (2006) found a positive relationship between study quality, reflected in the match between the unit of analyses and unit of assignment and treatment effectiveness.

Purpose

The purpose of this study was to add to the SCED literature investigating the effectiveness of peer tutoring on both academic and social performance for at-risk students and students with disabilities. By using a multilevel meta-analytic model, the overall average effectiveness was estimated over cases and over studies without losing information about individual studies or individual cases. The meta-analytic method accounted for the design type, trends in the data, as well as autocorrelation in a methodologically defensible way.

A secondary purpose of this study was to evaluate whether predictors at the second (between-case) or third (between-study) levels of the model would moderate treatment effects. If a large amount of between-case variance in treatment effects was found, we planned to evaluate predictors (i.e., age, gender, or disability type) that could explain between-case differences at the second level. If there was substantial between-study variance in treatment effects, we planned to evaluate whether a study-level predictors (i.e., study quality) moderated outcomes. The following research questions and hypotheses guided this study.

1. We investigated whether peer tutoring had a statistically and practically significant effect on academic outcomes and social outcomes across the identified SCED studies. We used a multivariate meta-analysis to evaluate the overall average effectiveness of peer tutoring across cases and across studies. We hypothesized peer tutoring would result in a significant effect on the academic and social outcomes for at-risk students and students with disabilities.
2. We investigated whether second-level or third-level predictors explained between-case or between-study

variability in the treatment effects. We evaluated whether age, gender, study quality, or disability type predicted variation in treatment outcomes. To investigate this last question, an interaction between the treatment effect and each of the moderators of interest was added to the model.

Method

Identification and Selection of Papers

Primary studies were retrieved using the following scientific databases: PsycINFO, Web of Science, MEDLINE, PubMed, and Educational Resources Information Clearinghouse (ERIC). Single-case experimental design studies investigating the effectiveness of peer tutoring on academic and/or social outcomes published between 1980 and 2014 were eligible for inclusion. The keywords used in the scientific databases are “single-case,” “single subject,” “N of 1,” “small N,” “multiple baseline design,” “alternating treatments design,” “reversal design,” “withdrawal design,” “interrupted time series” in combination with “peer tutoring,” “reciprocal peer tutoring,” “classwide peer tutoring,” “peers as tutors,” “peer-mediated instruction,” “peer-assisted learning,” and “across-age tutoring.” This initial search yielded a total of 216 unique published articles or unpublished dissertations that have the potential to be included in the meta-analysis. In addition, journals known to publish SCED articles were searched. Some of the journals did not allow to specify multiple keywords and as a consequence, only “single-case” was used as a keyword. For a full list of the journals that were hand-searched, see Online Appendix A.

Inclusion and exclusion criteria. The search was limited to work published in English, investigating peer tutoring as an intervention to increase academic performance and making use of a single-case design. Other peer-mediated interventions that targeted behavior directly (Kohler & Strain, 1990) were not included. All articles published in 2014 or earlier were eligible for inclusion. Articles not investigating peer tutoring as intervention, not evaluating social or academic outcomes, or only focusing on the tutor were excluded. Designs other than SCED, such as group comparison design studies, methodological papers (e.g., simulation studies and theoretical papers), or illustration papers were automatically excluded. Regarding the selection of SCED types, we focused on AB design, multiple-baseline design, reversal (or phase change) design, or alternating treatment designs (ATDs) with baselines. Other types of SCEDs, such as multiple probe designs, were excluded due to the fact that there are currently no recommendations that have been made about how to code the design matrix of these types of SCEDs (Moeyaert, Ugille, Ferron, Beretvas, & Van den Noortgate, 2014).

The fourth author read all the titles and the abstracts of the initial 216 retrieved articles and decided whether the article should be included or excluded based on the inclusion and exclusion criteria listed above. For 50% of the studies, the first author read the titles and abstracts. The percentage agreement for inclusion and exclusion was 78.45% between these two researchers, and 170 articles were excluded. In the case of disagreement, the assessors discussed the paper until they agreed on inclusion or exclusion. Of those 170 articles, 57 were excluded because the intervention did not involve peer tutoring. Two articles were excluded because peer tutoring was delivered within a multicomponent intervention and it was not possible to tease apart the unique effects of the peer tutoring. Several articles were excluded due to the type of design including 21 articles using group-comparison designs, 19 articles that were nonexperimental, and 12 articles that used multiple probe designs. There were eight articles that did not report baseline performance (ATD studies). These studies were excluded because the effectiveness of peer tutoring could not be evaluated in the same fashion as the other SCEDs in the analysis.

The analyses used in this study require raw data which were retrieved from time-series graphs. As a result, 26 articles without graphical presentations of data and two articles with unclear graphical presentations were excluded. In addition, we excluded two articles that did not include graphs for individual target students and six articles that only presented outcomes for the tutors (which was not an outcome of interest in this study).

A total of 10 articles were excluded because they did not include measures of academic or social outcomes, while three studies were excluded because participants were not receiving special education services, were not identified as having a disability, or were not identified as including students who were at risk. Finally, two articles were excluded because they were duplicates.

In sum, a total of 46 articles remained for inclusion in the meta-analysis. The first and fourth authors independently reviewed the remaining 46 articles to ensure they met the inclusion criteria. The percentage agreement for the final inclusion of the 46 articles between the two raters equaled 93.43%.

Study Coding

Studies were coded on 11 variables by the second, third, and fourth authors. A coding manual, containing descriptions of all relevant variables was created and can be found in Online Appendix B. Prior to the coding, the third and fourth authors familiarized themselves with the coding manual and test coded 10 articles that were not included in this study. A total of 20% of the articles were coded by both raters. Interobserver agreement between the two raters was

100%, with no discrepancies on any of the items. Each study was coded on the population of students served (e.g., special education classification), tutor and tutee's age and gender, intervention type (i.e., classwide peer tutoring, reciprocal peer tutoring, or nonreciprocal peer tutoring; Talbott et al., 2017), the dependent variable and measure used, the design type, study quality, and how researchers assessed the magnitude of any observed treatment effects. Below we describe the dependent measures and potential moderating variables in more detail.

Dependent variables. The focus of this study was on peer tutoring used to directly target academic skills and two dependent variables were of interest in the current study: academic performance and social-behavioral outcomes. Therefore, academic performance variables only included measures of academic skills (e.g., measures of reading, subject area knowledge, math skills). We did not classify related outcomes that supported academic skills (e.g., on-task behavior) in this category, because the focus of this study was on peer tutoring used to directly target academic skills. The specific target skill (e.g., reading fluency) and type of measure (e.g., words read correct per minute) were coded. Measures in this area included curriculum-based measures (early literacy and oral reading fluency), measures of word recognition and sight word reading, spelling performance (e.g., words spelled correctly, spelling errors), math computation and problem-solving measures (accuracy and fluency), idiom comprehension, and expressive language.

Social-behavioral outcomes also greatly varied in the included studies both in terms of skill and scale of measurement. Social-behavioral outcomes included social interactions with peers (e.g., frequency counts of initiations and responses, utterances, communicative strategy use), positive social behaviors (e.g., social engagement, participation, following directions), and negative social behaviors (e.g., aggression, negative verbalization). A variety of scales of measurement were also used including frequency counts, interval percentages, and observational rating scales. To account for these differences in academic and social-behavioral outcome measurements, a standardization procedure was applied prior to combining the effect sizes. Further information related to the standardization procedure can be found in Online Appendix C.

Independent variables. Four potential moderating variables were coded for use in the multilevel models. The age and gender for the tutor and the tutee were coded. Age was a continuous predictor, while gender was coded as a dichotomous predictor (0 if female and 1 if male).

Third, the study quality was coded according to the What Works Clearinghouse (WWC) Single-Case Design Standards (Kratochwill et al., 2010) requirements for

evidence standards and visual analysis. The WWC lists six features and four steps that need to be followed during quality analysis. Features used to analyze data include (a) level, (b) trend, (c) variability, (d) an immediacy of effect, (e) overlap, and (f) consistency. Researchers are required to follow four visual analysis steps: (a) observe a predictable baseline pattern, (b) examine data within each phase, (c) compare phases to determine if an effect is present, and (d) find a demonstrated effect to occur at least three times in the study in which three cases are included with at least five observations in the baseline level. If the graph met the WWC visual analysis requirements and the interobserver agreement is more than 20%, the study meets evidence standards. If the graph does not meet any WWC visual analysis requirements, the interobserver agreement is less than 20%, or the intervention is not systematically manipulated, the study does not meet evidence standards. Finally, a study is considered to meet evidence standards with reservations if interobserver agreement is more than 20% and the graph meets some but not all WWC requirements. Each of the studies was coded as Meets Standards (2), Meets Standards with Reservations (1), or Does Not Meet Standards (0).

Fourth, participants' disability status was coded. Student disability status was based on the primary disability and students identified as average or high-achieving, without a concomitant disability, were excluded from the study. Students who had deficits in an academic skill, who were identified as having delays in academic functioning, or were receiving tiered interventions, without any disability identified were classified as at risk/low achievement. We used the following categories to classify students identified as having a disability.

Autism spectrum disorders included students diagnosed with autism or pervasive developmental disorders—not otherwise specified. Behavioral disorders included students with emotional or behavioral disorders, emotional disturbance, or other health impairments when ADHD was the identified health impairment. This category also included students with *DSM* diagnoses such as oppositional defiant disorder and ADHD. Intellectual disability included students with intellectual disability or medical conditions with associated cognitive impairments (e.g., Down syndrome, Prader–Willi syndrome). Learning disabilities included students identified as having a specific learning disability.

Analysis

WebPlot Digitizer 2.0 was used to retrieve raw single-case data from the primary studies. This data retrieval software program appears to be the most reliable, valid, and user-friendly in addition to being free of charge (Moeyaert, Maggin, & Verkuilen, 2016). The basic procedures for extracting data follows a routine which includes (a)

importing the graph into the program, (c) defining the coordinate system, and (c) clicking on each data point (from the first observation to the last observation). Two columns of values are obtained: a column containing X-values (i.e., representing time, going from the beginning of the experiment to the end of the experiment) and a column with the Y-values (i.e., the dependent variable, representing the outcome scores per X-value). Finally, the researcher copies or exports the data to Microsoft Excel or a text file for secondary analyses. One researcher retrieved all of the raw data (total of 9,634 data points). A second researcher, who was blind to the study purpose, retrieved 20% of the data. An interobserver reliability of 89.21% was found. This meets the minimum required percent agreement of .80 (Hartmann, Barrios, & Wood, 2004).

Using the data collected from each of the included studies, the effectiveness of a peer tutoring intervention between and within studies could be determined using multilevel modeling. A multilevel model was created to determine whether the overall average intervention effect across the identified studies is statistically significant. This was done by examining the immediate treatment effect of the peer tutoring interventions and the effect of the interventions across time. Differences in immediate treatment effect of a peer tutoring intervention and the effect of the intervention over time between academic and social outcomes were also examined in addition to including age, gender, study quality, and disability moderators to explain the differences in variability between cases and studies. SAS Proc Mixed within SAS 9.4 (Copyright © 2015, SAS Institute Inc. SAS) was used to perform the multilevel meta-analysis. The Kenward–Roger method for estimating degrees of freedom was used as it contains a small sample bias correction that is recommended in single-case contexts (Ferron, Bell, Hess, Rendina-Gobioff, & Hibbard, 2009). For detailed descriptions of the equations and the parameters that were used, see Online Appendix C.

Results

Descriptive Statistics

A total of 46 SCED studies met all inclusion criteria and were included in the multilevel meta-analysis. There was an average of 3.60 participants per study ($Min = 1$, $Max = 12$, $Mdn = 3$, $SD = 1.84$). An average of 20.87 measurement occasions per participant ($Min = 5$, $Max = 77$, $Mdn = 18$, $SD = 10.79$) was found. Each study was coded on a variety of characteristics that were not included as moderators of intervention effectiveness. Multiple types of peer tutoring programs were used to address academic skills. We adopted the classification used by Talbott et al., (2017) to classify the types of programs. There were 13 studies of classwide peer tutoring, 12 studies of reciprocal

peer tutoring (which included PALS), and 21 studies of nonreciprocal peer tutoring included in the meta-analysis.

We coded multiple aspects of the study design. We differentiated between the three most common designs: 68.6% of the studies made use of a multiple-baseline design across participants, 26.2% used reversal designs, and 5.2% of studies implemented alternating treatment designs. We also coded how the study authors evaluated the magnitude of the treatment effects. Researchers used mean differences in 48.5% of the studies, 13.5% used PND, 1.4% used the Percentage of All Nonoverlapping Data (PAND), and the remaining used visual analysis alone.

Dependent variables. We are interested whether there is a differential effect of the peer tutoring intervention depending on the type of outcome. Academic performance was targeted in 70.7% of the cases. Academic variables included reading/literacy, language, and math skills. Researchers used measures of reading or literacy skills included measures of oral reading fluency, comprehension, reading accuracy, phonological/morphological awareness, and idiom comprehension. Measures of math skills included measures of basic math abilities and calculation.

In addition, 29.3% of the studies involved a social-behavioral outcome. For example, researchers measured social interactions with peers such as aggression during free play or positive behaviors during lunch time. Other measures included self-report items regarding motivation/attitude or social/behavioral outcomes.

Moderating variables. Four potential moderators (gender, age of the tutee, age of the tutor, study quality, and disability) were included in the analysis. Regarding gender, 33.0% of the tutees were female, 54.2% were male, and for 12.9% of the participants, gender was missing. Second, tutor gender was coded. Approximately, 20.1% of the tutors were female, 15.6% were male, and for 64.3% of the tutors, gender was missing. The age of the tutees ranged from 3 to 25, with a mean of 9.14 ($Mdn = 9.00$, $SD = 3.00$). Similar results were found for the tutor age: tutor age ranged from 4 to 18, with a mean age of 10.31 ($Mdn = 10.00$, $SD = 2.79$).

The third moderator was study quality. Approximately, 54.3% of studies met the WWC evidence standards, and 13.0% of studies met the WWC evidence standards with reservations. On the contrary, 32.6% of studies did not meet the WWC evidence standards. Common reasons for not meeting WWC standards were (a) less than five data points per phase was reported, (b) the study included fewer than three replications of the treatment effect, and (c) there were some noneffects.

Most of the participants were students with behavior disorders (36.59%), followed with children at risk/low achievement (29.1%), students with autism spectrum disorders (13.39%), students with learning disabilities (11.91%),

students with intellectual disability (6.86%), and students that were deaf (1.95%).

Inferential Statistics: Effect Size Estimation

Model 1. Model 1 predicts the immediate effect of peer tutoring and the linear time trend during the peer tutoring intervention for both academic and social outcomes (i.e., multivariate multilevel model). In other words, the treatment effect and trend during treatment for academic and social outcomes were analyzed. A statistically significant intervention effect for both academic, $\gamma_{200} = 4.18$, $t(28.8) = 2.41$, $p = .02$, and social outcomes, $\gamma_{200} = 1.84$, $t(16.1) = 3.94$, $p = .001$, was found. The time trend during the intervention phase was slightly positive and not statistically significant for academic outcomes, $\gamma_{300} = 0.04$, $t(19.1) = 1.09$, $p = .29$. The positive time trend indicates that the effect of the peer tutoring intervention increases across time. The reverse was found for social outcomes, $\gamma_{300} = -0.02$, $t(11.5) = -0.64$, $p = .53$.

The between-case variance of the treatment effect for academic outcomes ($\sigma_{u_b}^2 = 491.51$, $Z = 11.06$, $p < .0001$) and social outcomes ($\sigma_{u_i}^2 = 2.19$, $Z = 4.24$, $p < .0001$) was statistically significant, exhibiting the need to add predictors at the second level to explain some of the variability in treatment effect between cases (i.e., Model 2). No significant between-study variance was found for academic outcomes, whereas a significant amount of variance was found for social outcomes ($\sigma_{v_i}^2 = 3.27$, $Z = 2.27$, $p < .012$). Therefore, a third level moderator was added to the model as well (i.e., Model 2). An overview of the results can be found in Table 1.

Model 2. To explain variability in effectiveness of the peer tutoring intervention between studies and participants, moderators have been added. Therefore, gender, age, study quality, and disability moderators were added to the multilevel analysis in Model 2. In Table 1, the effectiveness of peer tutoring is given for the reference group, which are girls with learning disabilities, being at the average age ($M = 9.14$), and being part of a study that does not meet the WWC criteria for methodological rigor. Although not statistically significant, peer tutoring was associated with larger effects for older students (for both academic and social outcomes). For academic outcomes, the higher the study quality, the lower the effectiveness (for social outcomes, a very small positive effect was found, but this is negligible). The effect of gender was large and it appears that peer tutoring is less effective for male students as compared with female students. A larger difference is found for academic outcomes (-2.89) as compared with social outcomes (-0.56). Slightly different results were found for the effectiveness of peer tutoring as a function of the disability categories between academic and social outcomes. For both academic and social outcomes, peer tutoring

is less effective for children with intellectual disability. However, for academic outcomes, peer tutoring is most effective for children at risk/low achievement. For social outcomes, peer tutoring is most effective for children with autism spectrum disorders. As shown in Table 1, the between-case and between-study variance in the effectiveness of peer tutoring was reduced by adding these four moderators.

Discussion

The purpose of this study was to synthesize the direct effects of peer tutoring on academic outcomes and indirect effects of peer tutoring on social outcomes for at-risk students and students with disabilities in studies using SCEDs. We included 46 studies, which was more than two previous meta-analytic reviews of SCED research on peer tutoring including 26 and 20 studies, respectively (Bowman-Perrott et al., 2014; Bowman-Perrott et al., 2013). It should be noted that not only were peer-reviewed, published journal articles included in the meta-analysis but also nine unpublished dissertations. As publication bias is a common issue in the field of education research, the inclusion of unpublished works is especially important in decreasing the amount of bias in meta-analyses (Pigott, Valentine, Polanin, Williams, & Canada, 2013). Multilevel modeling was used to evaluate the effects of peer tutoring on the level and trend of students' academic and social outcomes. The use of multilevel modeling allowed for the investigation of between-phase, between-case, and between-study differences in the level and trend of student outcomes. That is, student-level or study-level variables could be entered into the models rather than relying on within-study aggregate values of potential moderators.

Overall, peer tutoring had a significant positive effects on academic and on social-behavioral outcomes. As hypothesized, peer tutoring had a larger impact on academic outcomes (i.e., direct effects) than on social-behavioral outcomes (i.e., indirect effects). Although not statistically significant, there was a slight positive trend for academic outcomes during intervention but the opposite was true for social-behavioral outcomes. This suggests that the positive, indirect impact of peer tutoring on social-behavioral outcomes may not continue to improve over time. There was substantial between-study and between-case variance in the effects of peer tutoring on academic and social outcomes necessitating the exploration of potential moderators of peer tutoring.

Moderators of Peer Tutoring

In an attempt to explain the between-case and between-study variability, we added case-specific and study-specific moderators to the three-level model. Although the moderators did not explain significant variation, these tests are

Table 1. Results From the Three-Level Multivariate Model.

Parameter	Parameter estimate (SE)	
	Model 1	Model 2
Fixed effects		
Intervention effects		
Intervention effect, academic	4.18* (1.74)	3.87 (2.04) ^a
Intervention effect, social	1.84* (0.47)	4.04* (1.04) ^a
Trend during treatment, academic	0.04 (0.03)	NA
Trend during treatment, social	-0.02 (0.03)	NA
Moderators		
Tutee Age × Intervention Effect, academic	NA	0.72 (0.67)
Tutee Age × Intervention Effect, social	NA	0.17 (0.17)
Study Quality × Intervention Effect, academic	NA	-0.30 (0.98)
Study Quality × Intervention Effect, social	NA	0.05 (0.65)
Tutee Gender × Intervention Effect, academic	NA	-2.89 (1.80)
Tutee Gender × Intervention Effect, social	NA	-0.56 (0.49)
Disability × Intervention Effect, academic	NA	
1. At risk/low achievement		3.65 (2.20)
2. Autism spectrum disorders		1.01 (2.82)
3. Behavioral disorders		-0.87 (2.15)
4. Deaf		NA
5. Intellectual disabilities		-2.48 (5.82)
Disability × Intervention Effect, social	NA	
1. At risk/low achievement		NA
2. Autism spectrum disorders		-0.77 (1.68)
3. Behavioral disorders		-2.57 (1.64)
4. Deaf		-2.85 (2.12)
5. Intellectual disabilities		-3.46 (2.04)
Random effects		
Between-study variance		
Intervention effect, academic	25.17 (19.5)	0
Intervention effect, social	3.27* (1.44)	0.16 (0.70)
Between-case variance		
Intervention effect, academic	491.51* (44.46)	0
Intervention effect, social	2.19* (0.51)	0

Note. Model 2 estimated the standardized mean difference between the intervention phase and the baseline phase and does not model changes in time trends due to the intervention. NA = not applicable; WWC = What Works Clearinghouse.

^aFor Model 2, the intervention effects are given for the reference group, which are girls with learning disabilities, being at the average age ($M = 9.14$), who participated in a study that does not meet the WWC criteria for methodological rigor.

* $p < .05$.

often underpowered (Borenstein, Hedges, Higgins, & Rothstein, 2009). Therefore, we interpret the moderator results here with substantial caution. Although peer tutoring had large positive effects on academic outcomes, the impact of peer tutoring was smaller in studies that met WWC criteria. The impact of study quality on social outcomes was almost negligible.

The gender of the target students appeared to have the potential impact on peer tutoring effects on academic and social-behavioral outcomes. Controlling for the other variables in the model, peer tutoring had larger direct and indirect effects for females than males. Peer tutoring had larger

effects for students who were aged 10 and older (recall the average age was 9.14 years old). The impact of age on academic effects was more pronounced than on social-behavioral outcomes.

Finally, the impact of peer tutoring differed based on student disability status. In comparison to students with learning disabilities, peer tutoring had a larger effect on academic skills for students who were identified as at risk or low-achieving. Perhaps unexpectedly, the impact of peer tutoring on academic outcomes was larger for students with autism spectrum disorders compared with students with learning disabilities, whereas students with

learning disabilities responded more positively to peer tutoring than students with behavior disorders or students with an intellectual disability.

In terms of indirect effects on social-behavioral outcomes, there were a few notable findings. First, researchers did not investigate the indirect effects of peer tutoring for students who were classified as at risk or low-achieving. Second, peer tutoring had the highest indirect effects for students with learning disabilities in comparison with the other disability categories. The difference between students with learning disabilities and students with autism spectrum disorders was less pronounced than for behavioral disorders, students who were deaf, or students with intellectual disability.

Interpretation of Findings in Context

Current standards for meta-analytic research suggest that results should be interpreted in context of meta-analytic reviews of similar interventions. Peer tutoring was associated with medium effect sizes ($d = 0.55$) in 14 meta-analytic reviews of between-group studies with 2,676 participants (Hattie, 2009). The average effect size in the current study was larger, which is often the case when comparing results from SCEDs to between-group designs (Shadish, Hedges, & Pustejovsky, 2014).

Previous studies by Bowman-Perrott and colleagues (Bowman-Perrott et al., 2014; Bowman-Perrott et al., 2013) may provide the most accurate comparison for the current study, as they investigated the effects of peer-mediated and peer tutoring interventions across multiple disability types and on academic and social-behavioral outcomes. Peer tutoring had a larger overall effect on academic outcomes ($k = 26$; $\text{Tau-U} = 0.75$; Bowman-Perrott et al., 2013) than indirect effects on social-behavioral outcomes ($k = 6$; $\text{Tau-U} = .43$; Bowman-Perrott et al., 2014). Similar to the previous studies by Bowman-Perrott and colleagues, we found that peer tutoring had larger direct and indirect effects for older students (see also Leung, 2015). The impact of participants' gender was not evaluated in previous meta-analytic reviews of SCED research, but the current results suggest it may be a moderator of peer tutoring effectiveness. Notably, we were unable to investigate if cross-gender pairings of students impacted peer tutoring effectiveness given that few studies reported the gender of the tutor.

Bowman-Perrott et al. (2014) and Bowman-Perrott et al. (2013) excluded studies that did not meet WWC standards for rigor, which is consistent with guidance from the WWC (Kratochwill et al., 2010). However, excluding these studies may result in a form of bias in the analyses. An alternative method is to evaluate whether study quality impacts treatment effects. Indeed, after controlling for age, gender, and disability status, study quality was negatively related to treatment effects. Thus, it seems important that consumers

of SCED research on peer tutoring be aware that treatment effects in studies that do not meet methodological rigor may be positively biased.

Finally, our results corroborate prior systematic reviews (e.g., McMaster et al., 2006) and meta-analyses (Bowman-Perrott et al., 2014) suggesting that disability status may moderate the effects of peer tutoring. Although the impact of peer tutoring may be mitigated based on disability severity (Al Otaiba & Fuchs, 2002), results from Bowman-Perrott et al. (2014) and this study suggest the relationship between peer tutoring effects and disability severity may not linear. For example, Bowman-Perrott et al. (2014) found that students with intellectual disability benefited the most from peer-mediated interventions that directly or indirectly targeted student behavior. Peer-mediated interventions had relatively similar effects for students with other health impairments (or ADHD), emotional-behavioral disorders, or learning disabilities, and lower effects for students with autism. However, Bowman-Perrott et al. (2014) appeared to ignore the nesting of students in calculating these effects. This may be one reason for the differences between Bowman-Perrott et al. (2014) and the current study described above.

Limitations and Future Directions

These results should be interpreted in the context of their limitations. Despite the increase in the number of included studies, we were unable to examine the differential impact of peer tutoring on various academic or social outcomes. Future research on the impact of peer tutoring on social or behavioral outcomes would allow for a more nuanced analysis between outcomes.

Another limitation is that meta-analysts must rely on what is reported in the primary studies to conduct the quantitative synthesis. In SCED studies, there is a tradition to graphically present the data. This allows the meta-analyst to retrieve the raw data from the graphs which is needed to calculate the effect size. An alternative is to use the effect sizes reported by the authors of the primary study to estimate the effects. Unfortunately, 26 articles we identified did not list the effect size or include a graphical presentation of the data. This reduced the number of studies we could include in the quantitative synthesis which may have biased the current results.

Finally, the studies included in the meta-analysis included a wide variety of social outcome measures. These include frequencies and percentages derived from observations and frequencies derived from rating scale systems such as the Multiple Option Observation System for Experimental Studies used in the Barton-Arwood (2003) study and the Peer Social Behavior Code of the Systematic Screening for Behavior Disorders used in the Plumer (2007) study. These differences in measurement of the social outcome variables make comparison of the outcomes more

difficult opposed to academic skill measures which covered a smaller scope of outcomes

Implications for Practice

Despite the limitations of this study, the results have some potential implications for practice. First, these results corroborate the broader literature identifying peer tutoring as an evidence-based practice for at-risk students and students with disabilities. Educators should be aware, however, that effects reported in studies with limited methodological rigor may be inflated. Second, these results corroborate previous evidence suggesting that older students may receive more benefit than younger students (Leung, 2015). Still, there are a number of studies suggesting that peer tutoring can be adopted for use in early grades (Fuchs & Fuchs, 2005; McMaster et al., 2006). Moreover, peer tutoring may be more effective for females than males (at least on average). Educators using peer tutoring to target male students' academic skills should closely monitor the students' progress to ensure it is having the desired impact. Third, peer tutoring may have the largest impact for students who were at risk or low-achieving, but educators could still consider the use of peer tutoring to directly address the academic and indirectly address social-behavioral outcomes for youth with disabilities (perhaps particularly for youth with learning disabilities and autism spectrum disorder). Other peer-mediated interventions (e.g., peer network, peer management) appear likely to have a larger effect on students' social-behavioral outcomes than peer tutoring that targets solely academic skills. Educators seeking to address the academic and social-behavioral outcomes of students with disabilities may wish to combine more than one type of peer-mediated interventions to concurrently improve the student's academic and social-behavioral skills.

Conclusion

There is a robust body of evidence from between-group designs supporting the use of peer tutoring. In this study, findings from 46 SCED studies provided further support for the use of peer tutoring. Peer tutoring had a larger effect on academic effects than social-behavioral outcomes. Practitioners can consider peer tutoring as an evidence-based approach for improving the level or trend in students' academic skills and level of students' social-behavioral outcomes. Future research is needed to determine the maintenance of the effects of peer tutoring and to establish whether the effects generalize to other outcomes.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported in part by the Institute of Education Sciences, U.S. Department of Education, through grant R305D150007. The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education.

ORCID iDs

David A. Klingbeil  <https://orcid.org/0000-0003-2571-4424>

Merve Turan  <https://orcid.org/0000-0003-2209-8910>

Supplemental Material

Supplemental material for this article is available online.

References

- Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention: A review of the literature. *Remedial and Special Education, 23*, 300–316.
- Barton-Arwood, S. M. (2003). *Reading instruction for elementary-age students with emotional and behavioral disorders: Academic and behavioral outcomes* (Unpublished doctoral dissertation). Vanderbilt University, Nashville, TN.
- Bene, K., Banda, D. R., & Brown, D. (2014). A meta-analysis of peer-mediated instructional arrangements and autism. *Review Journal of Autism and Developmental Disorders, 1*, 135–142. doi:10.1007/s40489-014-0014-9
- Borenstein, M., Hedges, L. V., Higgins, J., & 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). Chichester, UK: John Wiley. doi:10.1002/9780470743386.ch20
- Bowman-Perrott, L., Burke, M. D., Zhang, N., & Zaini, S. (2014). Direct and collateral effects of peer tutoring on social and behavioral outcomes: A meta-analysis of single-case research. *School Psychology Review, 43*, 260–285.
- Bowman-Perrott, L., Davis, H., Vannest, K., Williams, L., Greenwood, C., & Parker, R. (2013). Academic benefits of peer tutoring: A meta-analytic review of single-case research. *School Psychology Review, 42*, 39–55.
- Cohen, P. A., Kulik, J. A., & Kulik, C. L. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19*, 237–248. doi:10.3102/00028312019002237
- Cook, S. B., Scruggs, T. E., Mastropieri, M. A., & Casto, G. C. (1985). Handicapped students as tutors. *The Journal of Special Education, 19*, 483–492. doi:10.1177/002246698501900410
- Darney, D., Reinke, W. M., Herman, K. C., Stormont, M., & Jalongo, N. S. (2013). Children with co-occurring academic and behavior problems in first grade: Distal outcomes in twelfth grade. *Journal of School Psychology, 51*, 117–128. doi:10.1016/j.jsp.2012.09.005
- Dart, E. H., Collins, T. A., Klingbeil, D. A., & McKinley, L. E. (2014). Peer management interventions: A meta-analytic

- review of single-case research. *School Psychology Review*, 43, 367–384. doi:10.17105/SPR-14-0009.1
- Delquadri, J., Greenwood, C. R., Whorton, D., Carta, J. J., & Hall, V. (1986). Classwide peer tutoring. *Exceptional Children*, 52, 535–542.
- Devin-Sheehan, L., Feldman, R. S., & Allen, V. L. (1976). Research on children tutoring children: A critical review. *Review of Educational Research*, 46, 355–385. doi:10.3102/00346543046003355
- DuPaul, G. J., Ervin, R. A., Hook, C. L., & McGoey, K. E. (1998). Peer tutoring for children with attention deficit hyperactivity disorder: Effects on classroom behavior and academic performance. *Journal of Applied Behavior Analysis*, 31, 579–592. doi:10.1901/jaba.1998.31-579
- Elbaum, B., Vaughn, S., Tejero Hughes, M., & Watson Moody, S. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology*, 92, 605–619. doi:10.1037/0022-0663.92.4.605
- Ferron, J. M., Bell, B. A., Hess, M. F., Rendina-Gobioff, G., & Hibbard, S. T. (2009). Making treatment effect inferences from multiple-baseline data: The utility of multilevel modeling approaches. *Behavior Research Methods*, 41, 372–384. doi:10.3758/BRM.41.2.372
- Fuchs, D., & Fuchs, L. S. (2005). Peer assisted learning strategies: Promoting word recognition, fluency, and reading comprehension in young children. *The Journal of Special Education*, 39, 34–44. doi:10.1177/00224669050390010401
- Fuchs, D., Fuchs, L. S., Mathes, P. G., & Martinez, E. A. (2002). Preliminary evidence on the social standing of students with learning disabilities in PALS and no-PALS classrooms. *Learning Disabilities Research & Practice*, 17, 205–215. doi:10.1111/1540-5826.00046
- Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Education Research Journal*, 34, 174–206. doi:10.3102/00028312034001174
- Ginsburg-Block, M., & Fantuzzo, F. (1997). Reciprocal peer tutoring: An analysis of “Teacher” and “Student” interactions as a function of training and experience. *School Psychology Quarterly*, 12, 134–139. doi:10.1037/h0088955
- Ginsburg-Block, M. D., Rohrbeck, C. A., & Fantuzzo, J. W. (2006). A meta-analytic review of social, self-concept, and behavioral outcomes of peer-assisted learning. *Journal of Educational Psychology*, 98, 732–749. doi:10.1037/0022-0663.98.4.732
- Greenwood, C. R., Delquadri, J. C., & Hall, R.V. (1989). Longitudinal effects of classwide peer tutoring. *Journal of Educational Psychology*, 81, 371–383. doi:10.1037/0022-0663.81.3.371
- Hartmann, D. P., Barrios, B. A., & Wood, D. D. (2004). Principles of behavioral observation. In S. N. Haynes & E. M. Hieby (Eds.), *Comprehensive handbook of psychological assessment: Behavioral assessment* (pp. 108-127). Hoboken, NJ: Wiley.
- Hattie, J. A. C. (2009). *Visible learning*. New York, NY: Routledge.
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1, 39–65. doi:10.1002/jrsm.5
- Hinshaw, S. P. (1992). Academic underachievement, attention deficits, and aggression: Comorbidity and implications for intervention. *Journal of Consulting and Clinical Psychology*, 60, 893–903. doi:10.1037/0022-006X.60.6.893
- 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, 165–179. doi:10.1177/001440290507100203
- Jun, S., Ramirez, G., & Cumming, A. (2010). Tutoring adolescents in literacy: A meta-analysis. *McGill Journal of Education*, 45, 219–238. doi:10.7202/045605ar
- Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). New York, NY: Oxford University Press.
- King, K. R., Gonzales, C. R., & Reinke, W. M. (2018). Empirically derived subclasses of academic skill among children at risk for behavior problems and association with distal academic outcomes. *Journal of Emotional and Behavioral Disorders*. Advance online publication. doi:10.1177/1063426617754082
- Klingbeil, D. A., Van Norman, E. R., McLendon, K. E., Ross, S. G., & Begeny, J. C. (2019). Evaluating the tau-u effect size with curriculum-based measures of reading and the impact of measurement error. *Behavior Modification*, 43, 413–438. doi:10.1177/0145445518760174
- Kohler, F. W., & Strain, P. S. (1990). Peer-assisted interventions: Early promises, notable achievements, and future aspirations. *Clinical Psychology Review*, 10, 441–452. doi:10.1016/0272-7358(90)90047-E
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. Retrieved from <https://ies.ed.gov/ncee/wwc/Document/229>
- Kunsch, C. A., Jitendra, A. K., & Sood, S. (2007). The effects of peer-mediated instruction in mathematics for students with learning problems: A research synthesis. *Learning Disabilities Research & Practice*, 22, 1–12. doi:10.1111/j.1540-5826.2007.00226.x
- Leung, K. C. (2015). Preliminary empirical model of crucial determinants of best practice for peer tutoring on academic achievement. *Journal of Educational Psychology*, 107, 558–579. doi:10.1037/a0037698
- Maggin, D. D. (2015). Considering generality in the systematic review and meta-analysis of single-case research: A response to Hitchcock et al. *Journal of Behavioral Education*, 24, 470–482. doi:10.1007/s10864-015-9239-7
- Maheady, L., & Sainato, D. M. (1985). The effects of peer tutoring upon the social status and social interaction patterns of high and low status elementary students. *Education and Treatment of Children*, 8, 51–65.
- Mastropieri, M. A., Spencer, V., Scruggs, T. E., & Talbott, E. (2000). Students with disabilities as tutors: An updated research synthesis. *Advances in Learning and Behavioral Disabilities*, 14, 247–279.

- Mathes, P. G., & Fuchs, L. S. (1994). The efficacy of peer tutoring in reading for students with mild disabilities: A best-evidence synthesis. *School Psychology Review, 23*, 59–80.
- McDonnell, J., Mathot-Buckner, C., Thorson, N., & Fister, S. (2001). Supporting the inclusion of students with moderate and severe disabilities in junior high school general education classes: The effects of classwide peer tutoring, multi-element curriculum, and accommodations. *Education and Treatment of Children, 24*, 141–160.
- McMaster, K. L., Fuchs, D., & Fuchs, L. S. (2006). Research on peer-assisted learning strategies: The promise and limitations of peer-mediated instruction. *Reading & Writing Quarterly, 22*, 5–26. doi:10.1080/10573560500203491
- Moeyaert, M., Maggin, D., & Verkuilen, J. (2016). Reliability, validity, and usability of data extraction programs for single-case research designs. *Behavior Modification, 40*, 874–900. doi:10.1177/0145445516645763
- Moeyaert, M., Ugille, M., Ferron, J., Beretvas, S., & Van den Noortgate, W. (2014). The influence of the design matrix on treatment effect estimates in the quantitative analyses of single-case experimental design research. *Behavior Modification, 38*, 665–704. doi:10.1177/0145445514535243
- Moeyaert, M., Ugille, M., Ferron, J., Beretvas, S. N., & Van den Noortgate, W. (2013a). Three-level analysis of single-case experimental data: Empirical validation. *The Journal of Experimental Education, 82*, 1–21. doi:10.1080/00220973.2012.745470
- Moeyaert, M., Ugille, M., Ferron, J., Beretvas, S. N., & Van den Noortgate, W. (2013b). Three-level analysis of standardized single-case experimental data: A Monte Carlo simulation study. *Multivariate Behavior Research, 48*, 719–748. doi:10.1080/00273171.2013.816621
- O'Connor, M. A., & Daly, E. J., III. (2018). Selecting effective intervention strategies for escape-maintained academic-performance problems: Consider giving 'em a break! *Journal of School Psychology, 66*, 41–53. doi:10.1016/j.jsp.2017.09.003
- Okilwa, N. S., & Shelby, L. (2010). The effects of peer tutoring on academic performance of students with disabilities in grades 6 through 12: A synthesis of the literature. *Remedial and Special Education, 31*, 450–463. doi:10.1177/0741932509355991
- Parker, R. A., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy, 42*, 284–299. doi:10.1016/j.beth.2010.08.006
- Pigott, T. D., Valentine, J. C., Polanin, J. R., Williams, R. T., & Canada, D. D. (2013). Outcome-reporting bias in education research. *Educational Researcher, 42*, 424–432. doi:10.3102/0013189X13507104
- Plumer, P. J. (2007). *Using peers as intervention agents to improve the social behaviors of elementary-aged children with attention deficit hyperactivity disorder: Effects of a peer coaching package* (Doctoral dissertation). University of Massachusetts Amherst.
- Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology, 95*, 240–257. doi:10.1037/0022-0663.95.2.240
- Ryan, J. B., Reid, R., & Epstein, M. H. (2004). Peer-mediated intervention studies on academic achievement for students with EBD: A review. *Remedial and Special Education, 25*, 330–341. doi:10.1177/07419325040250060101
- Schaefer, J. M., Cannella-Malone, H. I., & Carter, E. W. (2016). The place of peers in peer-mediated interventions for students with intellectual disability. *Remedial and Special Education, 37*, 345–356. doi:10.1177/0741932516629220
- Shadish, W. R. (2014). Statistical analyses of single-case designs: The shape of things to come. *Current Directions in Psychological Science, 23*, 139–146. doi:10.1177/0963721414524773
- Shadish, W. R., Hedges, L. V., & Pustejovsky, J. E. (2014). Analysis and meta-analysis of single-case designs with a standardized mean difference statistic: A primer and applications. *Journal of School Psychology, 52*, 123–147.
- Spencer, V. G. (2006). Peer tutoring and students with emotional or behavioral disorders: A review of the literature. *Behavioral Disorders, 31*, 204–222. doi:10.1177/019874290603100206
- Stenhoff, D. M., & Lignugaris/Kraft, B. (2007). A review of the effects of peer tutoring on students with mild disabilities in secondary settings. *Exceptional Children, 74*, 8–30. doi:10.1177/001440290707400101
- Talbott, E., Trzaska, A., & Zurheide, J. L. (2017). A systematic review of peer tutoring interventions for students with disabilities. In M. T. Hughes & E. Talbott (Eds.), *The Wiley handbook of diversity in special education* (pp. 321–356). New York, NY: Wiley.
- Tarlow, K. R. (2017). An improved rank correlation effect size statistic for single-case designs: Baseline corrected tau. *Behavior Modification, 41*, 427–467.
- Tarlow, K. R. (2017). An improved rank correlation effect size statistic for single-case designs: Baseline corrected tau. *Behavior Modification, 41*, 427–467. doi:10.1177/0145445516676750
- Utley, C. A., & Mortweet, S. L. (1997). Peer-mediated instruction and interventions. *Focus on Exceptional Children, 29*(5), 1–23.
- Van den Noortgate, W., & Onghena, P. (2008). A multilevel meta-analysis of single-subject experimental design studies. *Evidence-Based Communication Assessment and Intervention, 2*, 142–151. doi:10.1080/17489530802505362
- Wexler, J., Reed, D. K., Pyle, N., Mitchell, M., & Barton, E. E. (2015). A synthesis of peer-mediated academic interventions for secondary struggling learners. *Journal of Learning Disabilities, 48*, 451–470. doi:10.1177/0022219413504997