



ELSEVIER

Contents lists available at ScienceDirect

Educational Research Review

journal homepage: www.elsevier.com/locate/edurev

Does co-teaching improve academic achievement for students with disabilities? A meta-analysis

Margaret E. King-Sears^{a,*}, Abraham Stefanidis^b, Sheri Berkeley^a, Vasilis Strogilos^c

^a George Mason University, Division of Special Education and disability Research, 4400 University Drive, Fairfax, VA, 22030, USA

^b St. John's University, Department of Management, Peter J. Tobin College of Business, 101 Astor Place, Room 230, New York, NY, 10003, USA

^c University of Southampton, University Road, Southampton Education School, Building 32, Southampton, SO17 1BJ, UK

ARTICLE INFO

Keywords:

Meta-analysis
Students with disabilities
Co-teaching
Achievement

ABSTRACT

Co-teaching consists of general and special education teachers who team to instruct students with and without disabilities in the same classroom. Additionally, some students with disabilities (SWD) receive instruction in special education settings. The purpose of this meta-analysis was to compare academic achievement of SWD in co-taught classes to achievement of SWD in special education settings. A comprehensive search of published and gray literature yielded 26 eligible studies with a total of 3,714 SWD. Results indicate a moderate effect size ($g = 0.47$) for students in co-taught settings compared to students in special education settings. Although the findings support SWD's academic achievement in co-taught classes, more information is needed to identify what distinguishes characteristics of the SWD in co-taught versus special education settings, as well as instruction in each setting. Factors related to results are described, including the lack of information about instructional practices and equivalency of students assigned to co-taught or special education settings. Implications for future researchers are provided, along with cautions about widespread placement of SWD in either setting without monitoring that effective instructional practices are in place, and that students receive specially designed instruction as stipulated on their Individualized Education Programs.

1. Introduction

Whether students with disabilities (SWD) achieve better academic outcomes in general or special education settings has been a long-standing debate (Cole et al., 2004; Zigmond, 2015). In several developed countries, including England (Department for Education, 2020) and the USA (U.S. Department of Education, 2021), most SWD are taught in general education settings for at least part of their school days. In the State of Indiana alone, more than 32% of special education teachers are engaged in co-teaching instruction (Indiana State Department of Education, 2021). However, the extant literature does not present conclusive results regarding SWD's learning in those classroom settings (Cook et al., 2017; Van Garderen et al., 2012; Weiss & Glaser, 2021).

Among the different types of special education settings, two are prevalent in general education schools. The first is the resource room, in which students receive specially designed instruction from special educators for less than half of the school day, and the second is the self-contained classroom, in which the students receive instruction from special educators for the majority of the school day

* Corresponding author.

E-mail addresses: mkingsea@gmu.edu (M.E. King-Sears), stefania@stjohns.edu (A. Stefanidis), sberkele@gmu.edu (S. Berkeley), V.Strogilos@soton.ac.uk (V. Strogilos).

<https://doi.org/10.1016/j.edurev.2021.100405>

Received 28 January 2021; Received in revised form 31 July 2021; Accepted 19 August 2021

1747-9388/© 2021

(Illinois State Board of Education, 2020). Different terms are used across countries to describe placements for SWD in general education settings. For example, in Israel, these placements are called special classes (Gindi, 2020), whereas in England, they are called specially resourced provisions or designated units (Department for Department of Education, 2015).

In recent years, several researchers have called for empirical evidence to assess the efficacy of varying student placements, especially of co-teaching (Cook et al., 2017). Although co-teaching is implemented in general education, in this study, we operationalize it as a service delivery model designed for SWD, who receive instruction alongside their same-age peers without disabilities from a team of general and special educators (Friend et al., 2010). The curriculum instruction provided when special and general education co-teachers combine content and pedagogical expertise is assumed to be richer and more useful for both SWD and students without disabilities (SWOD) (King-Sears et al., 2014; van Hoven et al., 2012). For example, Lochner et al. (2019) found that SWD and SWOD in co-taught classes exhibited higher levels of cognitive engagement, linked to learning, in co-taught compared to solo-taught general education classes. These findings illustrate Friend et al.'s (2010) observation that co-teaching can be responsive to the diversity of today's learners. However, SWD's learning in co-taught settings needs to be compared with that of SWD's in special education settings, in which all students have disabilities. Cook et al. (2017) note that special educators should not overly rely on co-teaching until there is more evidence regarding its effectiveness. To this end, empirical evidence about the effectiveness of co-teaching is needed (e.g., Bottge et al., 2018), especially because many international organizations encourage countries to educate SWD within general education (e.g., UNESCO, 2017; U.S. Department of Education, 2021).

1.1. Co-taught SWD's achievement

Fontana (2005) was among the first researchers who compared English and math final grades of students with learning disabilities (LD) in resource versus co-taught settings. She found that co-teaching improved student performance in English and math. On the contrary, comparing the reading and writing standardized test scores of ninth grade urban school students with LD, Murawski (2006) found no differences between special education and co-taught classes. Hang and Rabren (2009) compared special education settings prior to co-teaching with co-taught settings and found that co-teaching increased the academic performance scores of 58 SWD in one or more core content areas. Similarly, Brusca-Vega et al. (2011) observed improvement in achievement for co-taught SWD and SWOD on standardized science tests. More recently, Bottge et al. (2018), who compared the math performance of 65 SWD in urban and rural middle schools, reported that co-taught SWD achieved higher scores than SWD from resource rooms. Lastly, Cole et al. (2020) compared fourth- to eighth-grade SWD's reading and math state test results and found that SWD who spent at least 80% of their time in general education classes performed significantly better than those who spent less time.

1.2. Existing meta-analyses of Co-Teaching

Murawski and Swanson (2001) conducted the first meta-analysis of co-teaching in which they compared students' achievement between non-co-taught (solo-taught general classrooms or special classrooms) and co-taught settings. Their search yielded 89 articles, from which six studies had sufficient empirical data for inclusion. The studies were published from 1991 to 1998 and reported student achievement scores from standardized tests. According to the authors, two studies did not report students' disabilities, and most studies did not report how students were assigned to settings (i.e., group assignment). Additionally, the authors highlighted the importance of disaggregating data on students' achievement between co-taught and non-co-taught settings, emphasizing the included studies' lack of sufficient information to support this comparison. Furthermore, four out of the six studies did not separate results between SWD and SWOD, precluding measurement of the effect of co-teaching for each student group. Consequently, they combined data from solo-taught general education and special education settings and reported them by grade level. Overall, the authors found a medium positive effect size (.40) for co-teaching, with a quite wide range (0.08–0.95). Importantly, they raised concerns about the lack of fidelity data with regard to the co-teaching arrangements in all studies.

Based on a pool of 1,738 reports, in his unpublished meta-analysis, Khoury (2014) identified 20 studies from 1996 to 2013 that compared the achievement of SWD in co-taught to solo-taught general education or resource room special education settings. There was a positive effect size among the 13 dissertation reports ($g = 0.439$) which was larger than that of the seven studies from peer-reviewed journals ($g = -0.343$). Khoury found a significant effect ($g = 0.435$) in the academic outcomes of SWD comparing co-taught to resource classrooms. However, in his analysis, he did not take into consideration the type of student achievement test (e.g., norm-referenced or criterion referenced). Similar to Murawski and Swanson (2001), he reported several omissions in the primary studies, such as the lack of fidelity measures, identification of co-teaching models, or participants' characteristics.

Szumski et al. (2017) meta-analyzed 47 studies (extracted from an original pool of 94 studies) to examine how the presence of SWD in inclusive settings impacts the academic performance of SWOD. The examined settings consisted of solo-taught general education and co-taught classes. However, no results were provided specifically for co-taught classes, because they did not distinguish co-taught from inclusive data. Szumski et al. identified a weak but positive effect size for the academic achievement of SWOD in inclusive settings ($d = 0.12$). Search years ranged from 1980 to 2015, and all the included studies were confined to standardized tests. As in previous reviews (Khoury, 2014; Murawski & Swanson, 2001), the authors did not find adequate fidelity data in the primary studies. Notably, the authors' research purpose did not encompass identifying the effect of co-teaching for SWD, so they did not present any data and effect sizes pertaining specifically to SWD.

In a more recent meta-analysis, Losinski et al. (2019) examined the impact of co-teaching for secondary SWD's learning comparing co-taught with resource and solo-taught classes. Their search, which stopped in July 2017, yielded 578 studies, of which nine were el-

eligible for further analysis. The authors reported overall low quality of the included studies due to important omissions (e.g., fidelity implementation). As was the case with Murawski and Swanson's (2001) meta-analysis, co-teaching was compared with combinations of solo-taught general education and resource rooms. Losinski et al. (2019) identified a non-significant negative effect ($d = -0.012$) of co-teaching on the achievement of SWD. They also recommended that future researchers include in their meta-analyses gray literature (i.e., unpublished studies and publicly available dissertations, conference proceedings, government documents, white papers, working papers, reports, and evaluations), which has been largely overlooked, and compare the quality of unpublished studies with that of articles published in peer-reviewed journals.

However, to date, all three meta-analyses that focused on co-teaching for SWD (Khouri, 2014; Losinski et al., 2019; Murawski & Swanson, 2001) have not provided clear comparable data on SWD's academic achievement because they aggregated data between SWD and SWOD. In addition, none of the previous meta-analyses compared SWD's achievement per setting, but data from special education resource rooms or solo-taught general education classrooms were aggregated.

1.3. Purpose

The value of SWD accessing the general education curriculum in co-taught or special education settings rests on multiple considerations. The benefits derived from a social environment where students learn in the same co-taught classroom present substantial merit (Garrote et al., 2017). Conversely, some SWD hold that they receive more individualized instruction when taught in special education settings (Leafstedt et al., 2007). One of the most significant considerations regarding student learning is that of SWD's academic achievement in co-taught compared to that of their peers in special education settings.

Previous meta-analyses about academic achievement limit comparisons to SWD taught in secondary special education settings (Losinski et al., 2019), are dated (Murawski & Swanson, 2001), and focus on one type of special education setting (Khouri, 2014; Losinski et al., 2019). Moreover, they limit their search of gray literature sources solely to dissertations (Khouri, 2014; Losinski et al., 2019). However, a comprehensive gray literature search encompasses accessing databases for dissertations, reports, and conference proceedings in addition to alerting multiple international organizations and researchers to acquire unpublished manuscripts (Ferguson & Brannick, 2012; Paez, 2017). This comprehensive search has not been included in almost 60% of meta-analyses (Gage et al., 2017).

Based on our review of the existing meta-analyses regarding co-teaching SWD, an updated meta-analysis that expands inclusion criteria encompassing SWD from all grades, targeting SWD in more types of special education settings, and conducting a thorough gray literature search is necessary. Additionally, the examination of previously under-examined factors, such as grade levels and students' assignments to different classroom settings, which potentially impact students' academic achievement, is warranted. In this context, the current meta-analysis embarked on addressing the following research questions.

Research Question 1: Does SWD's achievement differ between special education classes and co-taught classes?

Research Question 2: Are there specific factors that make SWD's achievement differ between special education classes and co-taught classes?

2. Method

We complied with guidelines from Kugley et al. (2016) to acquire the universe of eligible studies across a 20-year period between 1999, the end year for Murawski and Swanson's (2001) meta-analysis, and 2019. We sought peer-reviewed articles and gray literature (e.g., dissertations; unpublished manuscripts; reports). Eight databases were used: Academic Search Complete; Conference Proceedings Citation Index-Science; ERIC; EBSCO; OmniFile Full Text Select [H.W. Wilson]; PsycINFO; Open Dissertations; and ProQuest Dissertations. Six keywords were searched: *co-teaching*, *coteach**, *co-operative teaching*, *cooperative teaching*, *collaborative teaching*, and *team teaching*. These keywords represent all the different terms that are used in the international literature to describe what we define as "co-teaching" in this meta-analysis. Consistent with gray literature procedures, we issued "calls to the field" (social media, listservs, and newsletters of multiple international education organizations [e.g., American Educational Research Association; Asia Pacific Educational Research Association; European Educational Research Association]) for unpublished studies (e.g., book chapters; reports; unpublished datasets); examined gray literature internet sites (e.g., <http://www.opengrey.eu/>); contacted leading co-teaching researchers; and conducted systematic hand searches of peer-reviewed journals (van Leeuwen & Janssen, 2019). Two teams of researchers and doctoral students determined study eligibility independent of each other to identify studies meeting eligibility criteria. They attained inter-rater reliability (IRR) of 91%, then resolved discrepancies to 100%.

2.1. Eligibility of studies

Six inclusion criteria were required for eligibility: (a) academic achievement comparisons were between K-12 SWD receiving instruction from general and special education co-teaching teams and K-12 SWD in special education settings receiving instruction from special educators; (b) disaggregated data specific to academic achievement for SWD per setting were available; (c) objective measures were used to measure student achievement; (d) each study was an original group research design with a comparison group; (e) the study was written in English; and (f) data to calculate effect sizes were available. Consistent with our research questions, we excluded studies focused on educators' teaming (i.e., grade-level teams; subject-level teams) or some variation of inclusion that did not identify co-teaching by special and general educators. Studies that did not use special education classes as a comparison setting or did not in-

clude disaggregated data for SWD were not eligible. We excluded studies that aggregated multiple grade levels' subjects and data (e.g., one mathematics score for SWD in grades 3 to 8), as well as use of subjective measures of academic achievement, such as report card grades that include attendance.

The universe of articles from the comprehensive search was 13,315, from which 5,470 were excluded because they were duplicates, covered irrelevant disciplines, or were otherwise clearly ineligible for our meta-analysis (see Fig. 1 PRISMA; Moher et al., 2009). Screening by title and abstract occurred for the remaining articles (7,833), from which a further 7,365 were excluded because they were not empirical studies in education or were practitioner articles about collaboration, teaming, or co-teaching. Full text review occurred for 468 articles to verify those focusing on student achievement (99), from which 26 met all inclusion criteria for this meta-analysis.

2.2. Coding conventions

We followed guidance from Harari et al. (2020) and Polanin et al. (2017) to develop the coding protocol and detailed codebook, which were used to extract data from the eligible studies. The information coded consisted of generic information (author/s, publication source, year); descriptors of the school system and school (e.g., urban/public) as well as the co-teaching (e.g., models used; de-

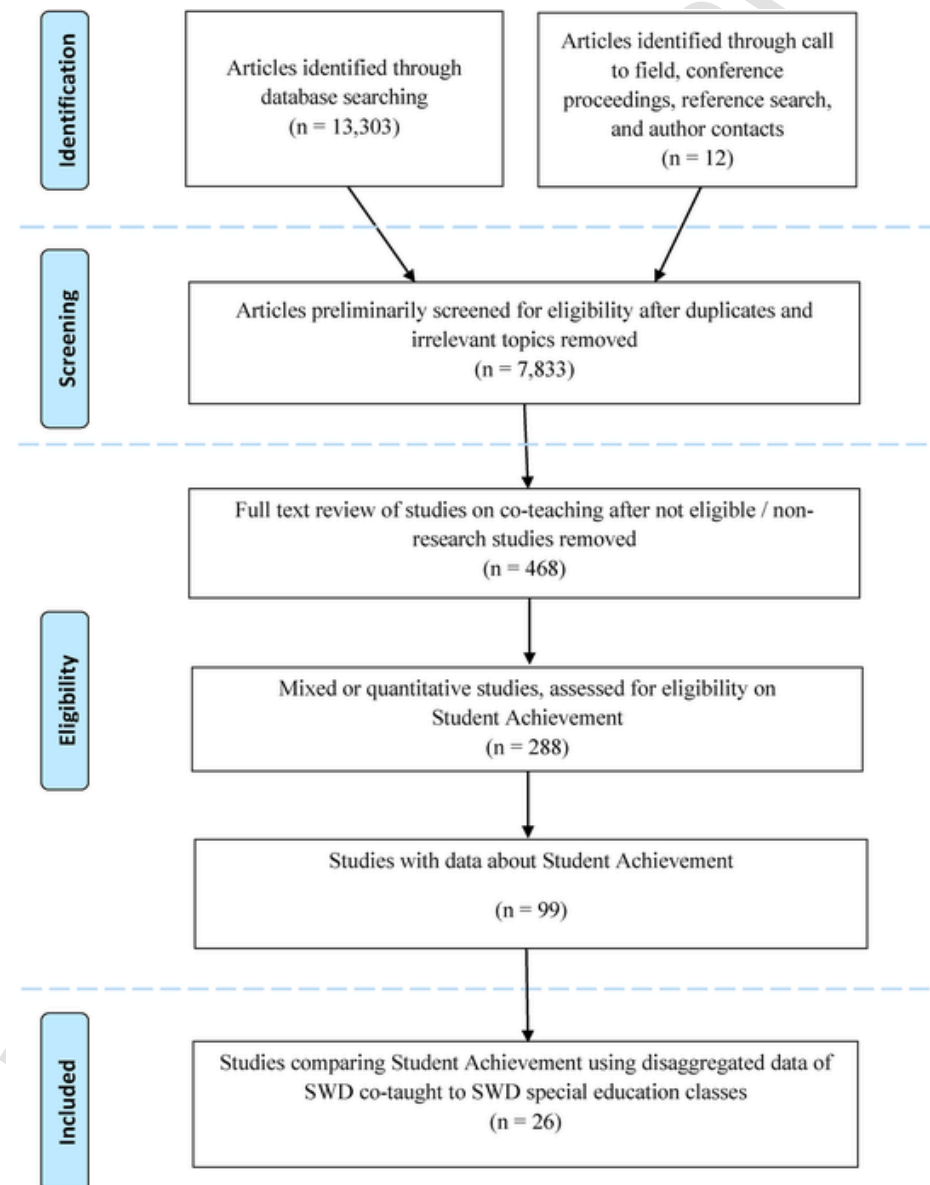


Fig. 1. PRISMA Flow Diagram Co-Teaching Studies on Student Achievement for SWD.

scription about instruction) settings; identification about attrition and fidelity; and details about the measures for student achievement (type of measurement with corresponding reliability/validity information and what content/skill was measured).

Further, we coded for the number of SWD in the co-teaching and special education settings as well as students' disability labels, grade levels, socioeconomic status, and other demographics, such as ethnicity and gender. The number of teachers and corresponding demographic information (e.g., ethnicity, gender), teachers' certification and teaching experience (e.g., years teaching/co-teaching), preparation for co-teaching (e.g., amount/content), and descriptions of co-teaching roles were also coded.

2.3. Study quality

A total of 33 items indicating study quality were rated using a binary system (yes/no). See [Table 1](#) for specific items per study and [Table 2](#) for items summed across studies. All items were derived from researchers' recommendations for study quality examinations (e.g., [Appelbaum et al., 2018](#); [Cook et al., 2014](#); [Gersten et al., 2005](#); [Talbot et al., 2017](#); [Valentine & Cooper, 2008](#)). Four researchers (two teams) received training to code all items. Post-coding, 96% IRR was attained, followed by resolution to 100% agreement. Certain items were targeted as moderators, such as group assignment and type of measurement.

2.4. Effect size calculation and analyses

From the studies that met the inclusion criteria, we extracted sample sizes, mean values, standard deviations, *t* values, and *F* values. Inter-rater reliability for the coding of the data entered was 97%, then resolved to 100%. We first compared scores of academic achievement (test outcomes) between students from co-taught classrooms and students from special education classrooms, and we calculated Cohen's *d* (standardized mean differences) corrected for sample size bias ([Hunter & Schmidt, 2004](#)). To prevent overestimation bias that may derive from smaller samples, we consequently transformed Cohen's *d* into Hedge's *g*. Furthermore, we calculated confidence interval limits, standard errors, and weights for the effect sizes (see forest plot: [Fig. 2](#)). We further acknowledged a certain extent of heterogeneity among the included studies, and, thus, we applied a random effects model to pool our studies, accounting for intra-study and inter-study variability ([Lipsey & Wilson, 2000](#); [Suurmond et al., 2017](#)).

3. Results

3.1. Study characteristics

From the 26 eligible studies, four were from peer-reviewed journals, including a dissertation ([Weichel, 2001](#)) that was subsequently published ([Murawski, 2006](#)), and 22 were dissertations. Although our search dates were 1999–2019, the eligible studies ranged from 2001 to 2018 (see [Table 1](#)). Based on the reported findings of these studies, we derived 89 outcome points (i.e., measurements). In total, our analysis included 7,415 academic achievement scores of SWD.

Scores were derived from school settings that were urban (29.2%), suburban (20.2%), rural (10.2%), or combinations of them (5.6%). For 34.8% of the academic achievement scores, the school setting was not specified.

All but one study reported SWD were receiving instruction either in co-taught, resource, or self-contained special education settings located at general education schools. One study noted the self-contained class was located at a "not traditional school," which may have been a general education or special school site ([Doran, 2008](#), p. 44). For special education settings, 50.6% were resource rooms, 30.3% were self-contained, 14.6% reported a mix of resource and self-contained, whereas 4.5% were not reported.

Schools' socio-economic status was reported in 42.3% of the studies. Although we searched internationally, only one study had been conducted in Belgium, with all other studies occurring in the United States, primarily in eastern states from New York to Georgia. Most studies occurred in southeastern states (12 in Georgia), two in Virginia, and one each in Arizona, California, Delaware, Florida, Kentucky, North Carolina, New Jersey, New York, South Carolina, Tennessee, and Texas.

3.1.1. Participant characteristics

From all studies, nine (34.6%) used more generic descriptors to identify SWD, such as students with special needs. The remaining studies (17 out of 26; 65.4%) identified students' disability categories, and one study provided detail about the special education eligibility procedures. Eight studies included only students with specific learning disabilities (LD). Almost all 17 studies that identified disabilities included students with LD (94.1%). Multiple disability categories (e.g., emotional disturbance or autism) were represented across the other nine studies. Thirty-nine percent (38.5%) of the studies reported students' low socioeconomic status. Students were in high school (29.2%), middle school (42.7%), and elementary school (28.1%) grades.

Similar demographic information was coded for co-teachers, as well as other relevant information (e.g., professional certification). Two studies identified co-teachers' gender, one study noted ages, and ethnicity was not reported in any of the studies. About one-fifth (19.2%) of studies noted co-teachers had received professional development on co-teaching. Certification (19.2%) and years of teaching (23.1%) were most frequently reported. Few studies (7.6%) reported years of co-teaching or years of co-teaching with the same partner.

Table 1
Summary of 26 Studies in Meta-Analysis.

Author/s	N SWD Co- Teach	N SWD Special Education Setting	Type of Special Education Setting	Grade Levels	Subject Areas	Measures	Types of Measurement	Co-Teaching Description
Adams (2014)	75	75	Resource	9, 10, 11, 12	Math	Georgia End of Course Test	Criterion-Referenced, State Test	Actions: N/A, Models: Team teaching, Fidelity: N/A*
Andrews-Tobo (2009)	24	15	Self-Contained	7, 8	Math, Reading	Georgia Criterion Referenced Competency Test	Criterion-Referenced, State Test	Actions: N/A, Models: N/A, Fidelity: N/A
Beachum (2016)	309	145	Resource	6	Math, Reading	North Carolina End of Grade	Criterion-Referenced, State Test	Actions: Behavior management; Small group or individual instruction, Models: One teach one observe or assist; Alternative teaching, Fidelity: N/A
Beam (2005)	32	34	Combination	3, 5	Math, Reading	Virginia Standards of Learning	Criterion-Referenced, State Test	Actions: N/A, Models: N/A, Fidelity: N/A
Bottge et al. (2018)	56	146	Resource	6, 7, 8	Math	Fractions Computation Test	Criterion-Referenced Test	Actions: High or low support roles; Multi-media and hands-on math instruction, Models: One-teach, one-assist; Team teaching, Fidelity: Identified
Castro (2007)	70	14	Combination	1, 2	Math, Reading	Terra Nova	Norm-Referenced Test	Actions: N/A, Models: N/A, Fidelity: N/A
Clements (2012)	89	45	Not Reported	7	Math, Reading	Tennessee Comprehensive Assessment Program	Criterion-Referenced, State Test	Actions: N/A, Models: N/A, Fidelity: N/A
Doran (2008)	300	300	Self-Contained	9, 10, 11, 12	Math, Reading, Science	End of Course Test	Criterion-Referenced, State Test	Actions: N/A, Models: N/A, Fidelity: N/A
Franklin (2015)	174	174	Combination	6, 7, 8	Math	Georgia Criterion Referenced Competency Test	Criterion-Referenced, State Test	Actions: N/A, Models: N/A, Fidelity: N/A
Gerlach (2017)	35	34	Resource	4	Reading	State of Texas Assessment of Academic Readiness	Criterion-Referenced, State Test	Actions: N/A, Models: N/A, Fidelity: N/A
Holmes (2018)	50	46	Combination	8	Math, Reading	Star Assessment	Norm-Referenced, State Test	Actions: N/A, Models: N/A, Fidelity: N/A
King (2015)	58	63	Resource	6, 7, 8	Math, Reading	Palmetto Assessment State Standards	Criterion-Referenced, State Test	Actions: N/A, Models: N/A, Fidelity: N/A
Laffitte (2012)	25	31	Resource	3, 4, 5	Reading	Galileo Criterion Referenced Assessment	Criterion-Referenced Test	Actions: N/A, Models: N/A, Fidelity: N/A
McCullough (2008)	24	19	Resource	8	Math	Delaware Student Testing Program	Criterion-Referenced, State Test	Actions: N/A, Models: N/A, Fidelity: N/A
Mote (2010)	80	77	Resource	6, 7, 8	Reading	Georgia Criterion Referenced Competency Test, Lexile Framework	Criterion-Referenced, State Test	Actions: Co-plan; re-teach; small group instruction, Models: One teach, one observe; Parallel, Fidelity: N/A
Nash-Aurand (2013)	74	71	Resource	9, 10, 11, 12	Math	Georgia Mathematics II End of Course Test	Criterion-Referenced, State Test	Actions: N/A, Models: N/A, Fidelity: N/A
Packard et al. (2011)	9	5	Resource	9	Reading	End of Course Test	Criterion-Referenced, State Test	Actions: Co-planning, Models: Parallel; Alternative, Fidelity: N/A
Popp (2001)	22	20	Resource	3, 5	Math, Reading, Social Studies, Science, Writing, Technology	Stanford-9 TA	Norm-Referenced, State Test	Actions: N/A, Models: N/A, Fidelity: N/A

(continued on next page)

Table 1 (continued)

Author/s	N SWD Co- Teach	N SWD Special Education Setting	Type of Special Education Setting	Grade Levels	Subject Areas	Measures	Types of Measurement	Co-Teaching Description
Powell (2007)	37	36	Resource	8	Math, Reading	Georgia Criterion Referenced Competency Test	Criterion- Referenced, State Test	Actions: N/A, Models: N/A, Fidelity: N/A
Reese (2017)	180	24	Resource	6, 7, 8	Math, Reading, Social Studies, Science, English Language Arts	Georgia Criterion Referenced Competency Test	Criterion- Referenced, State Test	Actions: N/A, Models: N/A, Fidelity: N/A
Scalise (2015)	31	34	Self-Contained	3, 4, 5, 6, 7, 8	Math, Reading	Northwest Evaluation Association Measures of Academic Progress	Norm- Referenced, State Test	Actions: N/A, Models: N/A, Fidelity: N/A
Shaw (2002)	66	135	Resource	5	Math, Reading	Stanford Achievement Test	Norm- Referenced Test	Actions: N/A, Models: N/A, Fidelity: N/A
Tremblay (2013)	58	105	Self-Contained	1, 2	Reading	Observation Survey TEDI-MATH	Norm- Referenced Test	Actions: N/A, Models: N/A, Fidelity: N/A
Walker Harris (2009)	53	54	Self-Contained	7, 8	Math, Reading	Georgia Criterion Referenced Competency Test	Criterion- Referenced, State Test	Actions: N/A, Models: N/A, Fidelity: N/A
Weichel (2001)	12	14	Resource	9	Math, Reading, Writing	Wide Range Achievement Test, Test of Reading Comprehension, Test of Written Language	Norm- Referenced Test	Actions: Co-planning; Facilitating student activity/learning/instruction, Models: One teach, one support; Team teaching, Fidelity: Identified
Williams (2012)	32	23	Resource	7, 11	Math, Reading	Georgia Criterion Referenced Competency Test	Criterion- Referenced, State Test	Actions: N/A, Models: N/A, Fidelity: N/A

*N/A: Not applicable.

3.1.2. Co-teaching descriptions

We reported how teachers implemented co-teaching as described by the researchers' observations. Five studies (19.2%) noted co-teaching actions (e.g., hands-on math instruction; re-teaching), and two (7.7%) of these studies identified fidelity. Specific co-teaching models were identified in six studies (23.1%). Most researchers (80.8%) described co-teaching by noting that instructional teaming between general and special educators occurred, with no further description.

3.1.3. Types of measures and content areas

Most studies (88.5%) used one measure, whereas three studies (11.5%) used two or more measures, yielding a total of 30 measures across all studies. Measurement types included norm-referenced (33.7%) and criterion-referenced (66.3%) assessments. Sixty-nine percent (69.2%) of the studies reported state test results. Ten studies (38.5%) focused on one content area, with 16 (61.5%) focused on two or more content areas. The tested content areas included: reading/language arts (49.4%), math (39.3%), and other (e.g., science; social studies; technology) (11.2%).

3.1.4. Categories for study quality ratings

The range of study quality ratings was from 41% to 76%. There is not widespread guidance for categorizing quality (Austin et al., 2019; Marusic et al., 2020), and some suggest deferring to researchers' judgments (Brown & Sutton, 2010). Subsequently, we followed guidance from those who adopted 50% and 80% as cut scores to sort studies into lower, moderate, and higher quality studies (Burke et al., 2020; Cook et al., 2014; Royer et al., 2017).

3.2. Co-teaching effects

We identified a medium positive combined effect for co-teaching ($g = 0.47$; $SE = 0.06$; 95% C.I. [0.35, 0.58]; $p = 0.00$), which indicated that the collective academic achievement of students from co-taught classrooms was moderately better than that of students from special education classrooms. Our forest plot (see Fig. 2) showed that the meta-analyzed studies included 76 positive effects and 13 negative effects for co-teaching. I^2 indicated that the proportion of observed variance which reflects real differences in effect size was 82.75%. This proportion was large enough to warrant subgroup analyses due to heterogeneity (Borenstein et al., 2011). The value of dispersion of true effect sizes between studies τ^2 was 0.28.

Based on subgroups' analyses, effects differed depending on students' grade level ($p = 0.032$). Compared to students from elementary schools ($g = 0.25$; 95% C.I. [0.05, 0.44]; weight = 30.37%; $I^2 = 61.81\%$; $\tau^2 = 0.13$), high school ($g = 0.52$; 95% C.I.

Table 2
Quality of Study Items Averaged for 26 Studies *.

	Quality Item	% Met Criteria
Context	Are the settings identified as co-taught or special education?	100%
	Is the school identified as urban, suburban, or rural?	69%
	Is the school identified as public or private?	62%
	Is the geographic location identified?	100%
	Is the curriculum identified?	100%
Students	Is the socioeconomic status of the school system identified?	42%
	Are students' grade levels identified?	100%
	Are students' genders identified?	42%
	Are students' ethnicity identified?	42%
	Are students with low SES reported?	65%
	Are students' disabilities label identified?	65%
	Is the process for students' eligibility for IEPs identified?	4%
Co-Teachers	Are co-teaching teams comprised of general and special educators?	100%
	Is certification for co-teachers identified?	19%
	Is ethnicity for co-teachers identified?	0%
	Is gender for co-teachers identified?	9%
Co-Teaching Description	Was training on co-teaching reported?	19%
	Are actions about co-teaching identified?	19%
	Are co-teaching models used in the study identified?	23%
Student Achievement Measures	Is fidelity about co-teaching identified?	8%
	Are the student achievement outcomes socially important?	100%
Internal Validity	Is the measure's name, purpose, and construct identified?	100%
	Are data per student achievement measure provided?	100%
	Are the frequency and timing of measures appropriate?	100%
	Is the type of measurement identified?	100%
	Is the specific content of the measure identified?	100%
	Is evidence of adequate reliability per measure provided?	70%
	Is evidence of adequate validity per measure provided?	64%
Internal Validity	Is student access to other instructional groups limited or eliminated?	100%
	Did students' assignment include attention toward group's similarity?	15%
	Is attrition identified?	12%
	From studies reporting attrition, is it low?	100%

* Detailed quality scores per study can be obtained from the corresponding author.

[0.41, 0.63]; weight = 40.65%; $I^2 = 26.50\%$; $\tau^2 = 0.02$) and middle school students ($g = 0.56$; 95% C.I. [0.35, 0.76]; weight = 28.99%; $I^2 = 90.62\%$; $\tau^2 = 0.47$) benefitted the most from co-taught classrooms. For elementary school students, the effect size was comparably lower.

In terms of content area tested in the exam, effects were similar for 'language arts/literacy' ($g = 0.60$; 95% C.I. [0.46, 0.74]; weight = 42.69%; $I^2 = 67.97\%$; $\tau^2 = 0.15$) and 'math' ($g = 0.42$; 95% C.I. [0.25, 0.59]; weight = 39.95%; $I^2 = 85.98\%$; $\tau^2 = 0.31$), but for 'other' subjects, the analysis did not identify an effect ($g = 0.08$; 95% C.I. [-0.35, 0.51]; weight = 17.36%; $I^2 = 88.33\%$; $\tau^2 = 0.45$). Co-teaching had a medium-sized effect on the academic achievement in 'language arts/literacy' and 'math,' whereas it did not have an effect in the 'other' category ($p = 0.052$).

Varying medium positive effects were observed regarding the type of tests that the students took ($p = 0.049$). For norm-referenced tests, co-teaching had a less positive effect on students' performance ($g = 0.32$; 95% C.I. [0.16, 0.47]; weight = 49.32%; $I^2 = 61.72\%$; $\tau^2 = 0.10$) than it had for criterion-referenced tests ($g = 0.55$; 95% C.I. [0.40, 0.69]; weight = 50.68%; $I^2 = 86.45\%$; $\tau^2 = 0.36$).

However, the subgroup analysis revealed that there were no significant differences ($p = 0.919$) in the academic achievement of students depending on whether they took a state exam ($g = 0.48$; 95% C.I. [0.34, 0.62]; weight = 63.17%; $I^2 = 85.06\%$; $\tau^2 = 0.33$) or a non-state exam ($g = 0.43$; 95% C.I. [0.25, 0.61]; weight = 36.83%; $I^2 = 69.20\%$; $\tau^2 = 0.13$).

In addition, we performed subgroup analysis considering the reported setting of the schools that the students attended. We identified variance in the academic achievement of students depending on school setting, yet these differences were not statistically significant ($p = 0.059$). Specifically, the effect size of co-teaching was stronger for rural schools ($g = 0.64$; 95% C.I. [0.38, 0.90]; weight = 25.20%; $I^2 = 71.55\%$; $\tau^2 = 0.11$) and for urban schools ($g = 0.57$; 95% C.I. [0.41, 0.72]; weight = 33.32%; $I^2 = 58.75\%$; $\tau^2 = 0.10$), whereas it was smaller for suburban schools ($g = 0.28$; 95% C.I. [-0.01, 0.57]; weight = 23.18%; $I^2 = 89.64\%$; $\tau^2 = 0.57$). Also, the effect size of co-teaching for studies that included achievement outcomes from a combination of urban, suburban, or rural school settings was close to zero ($g = 0.12$; 95% C.I. [-0.25, 0.49]; weight = 18.29%; $I^2 = 71.92\%$; $\tau^2 = 0.12$).

We also attempted to extract data concerning whether the schools reported in the studies were public or private. However, eight studies did not report the school settings where the academic achievement outcomes were retrieved from. We, thus, considered that differences per school setting could have been strengthened or weakened, if all school settings had been identified. Moreover, the au-

Forest Plot (Part I)

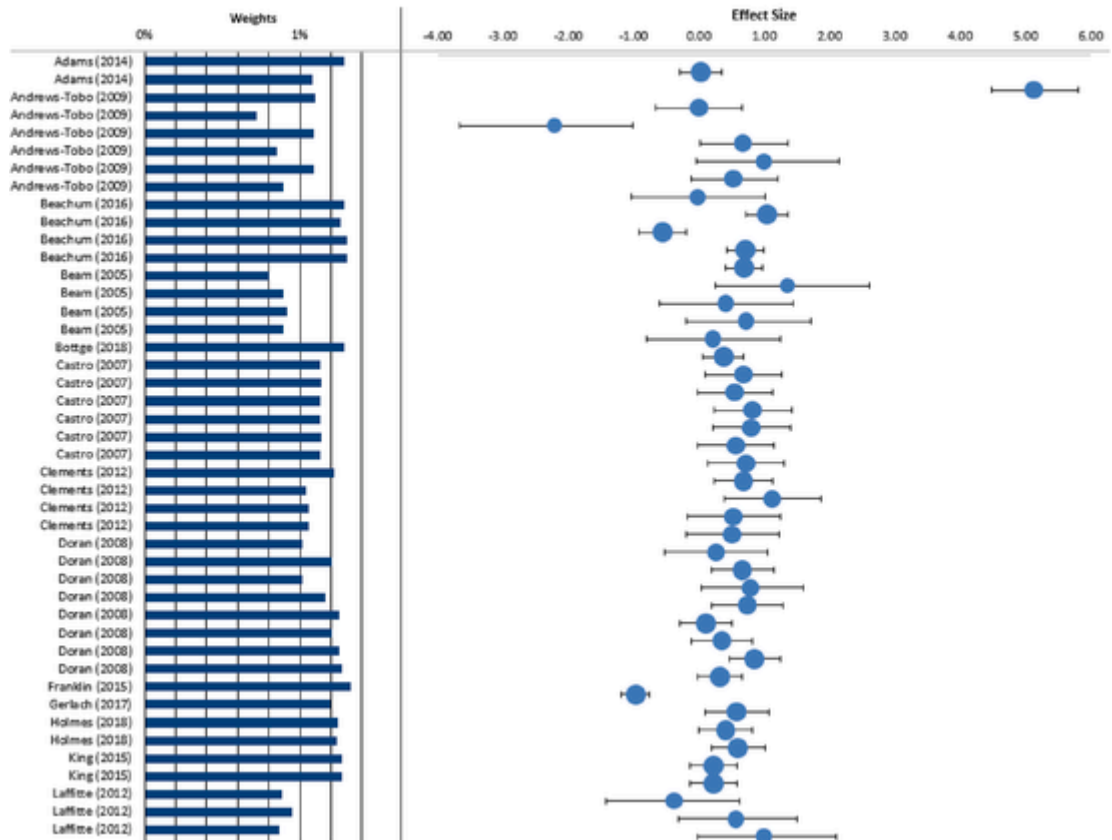


Fig. 2a. Forest plot (part I).

thors of the studies specified the type of schools only for 60% of the students' achievement results. Yet, of all 60% of the students' achievement results reported, 100% of schools were public, so we could not further perform a subgroup analysis.

No significant differential effects of student achievement were observed depending on whether the control group of students was taught in resource rooms or in self-contained classes ($p = 0.693$). When comparing the co-taught students' achievement with that of students from resource rooms ($g = 0.41$; 95% C.I. [0.23, 0.59]; weight = 37.26%; $I^2 = 83.99%$; $\tau^2 = 0.27$), the effect was comparably lower than when comparing the co-taught students' achievement with that of students taught in self-contained classes ($g = 0.51$; 95% C.I. [0.36, 0.67]; weight = 48.78%; $I^2 = 61.02%$; $\tau^2 = 0.10$) or with that of students taught in a combination of resource rooms and self-contained classes ($g = 0.50$; 95% C.I. [0.21, 0.80]; weight = 13.96%; $I^2 = 90.69%$; $\tau^2 = 0.68$). One study did not specify what was the type of the non-co-taught special education class, so it was not included in the analysis.

We identified a significant difference in the effect size of co-teaching derived from studies that adopted a convenience sampling research methodology versus the effect size of co-teaching derived from those studies that adopted randomization or otherwise established equivalence (e.g., matched pairs) per condition ($p = 0.022$). The positive effect of co-teaching from studies that adopted a convenience sampling research methodology ($g = 0.51$; 95% C.I. [0.38, 0.64]; weight = 52.46%; $I^2 = 85.15%$; $\tau^2 = 0.31$) was significantly stronger than the effect of co-teaching from studies that adopted a non-convenience sampling research methodology ($g = 0.24$; 95% C.I. [0.06, 0.41]; weight = 47.54%; $I^2 = 34.08%$; $\tau^2 = 0.05$).

We examined variations in the effect sizes based on the studies' publication year, with division at the midpoint year of 2009. No significant variations were observed ($p = 0.655$). For studies published until 2009, the medium positive effect size of co-taught classrooms ($g = 0.43$; 95% C.I. [0.29, 0.58]; weight = 57.81%; $I^2 = 62.48%$; $\tau^2 = 0.13$) was similar to that reported in studies published after 2010 ($g = 0.49$; 95% C.I. [0.32, 0.66]; weight = 42.19%; $I^2 = 88.69%$; $\tau^2 = 0.37$).

We assessed whether effects of co-teaching varied depending on the quality of the published studies, yet no significant differences were identified ($p = 0.693$). The effect of co-teaching on academic achievement for studies with 'quality < 50%' ($g = 0.52$; 95% C.I. [0.19, 0.85]; weight = 11.39%; $I^2 = 86.76%$; $\tau^2 = 0.41$) was comparable to the effect of co-teaching for studies with '50% < quality < 80%' ($g = 0.45$; 95% C.I. [0.34, 0.57]; weight = 88.61%; $I^2 = 81.70%$; $\tau^2 = 0.26$). Overall, the mean quality score of the included studies was 57.46% ($SD = 7.95%$). We did not identify a correlation between the effect sizes and the quality scores. Moreover, we did not identify a correlation between study quality and publication year of each study.

Forest Plot (Part II)

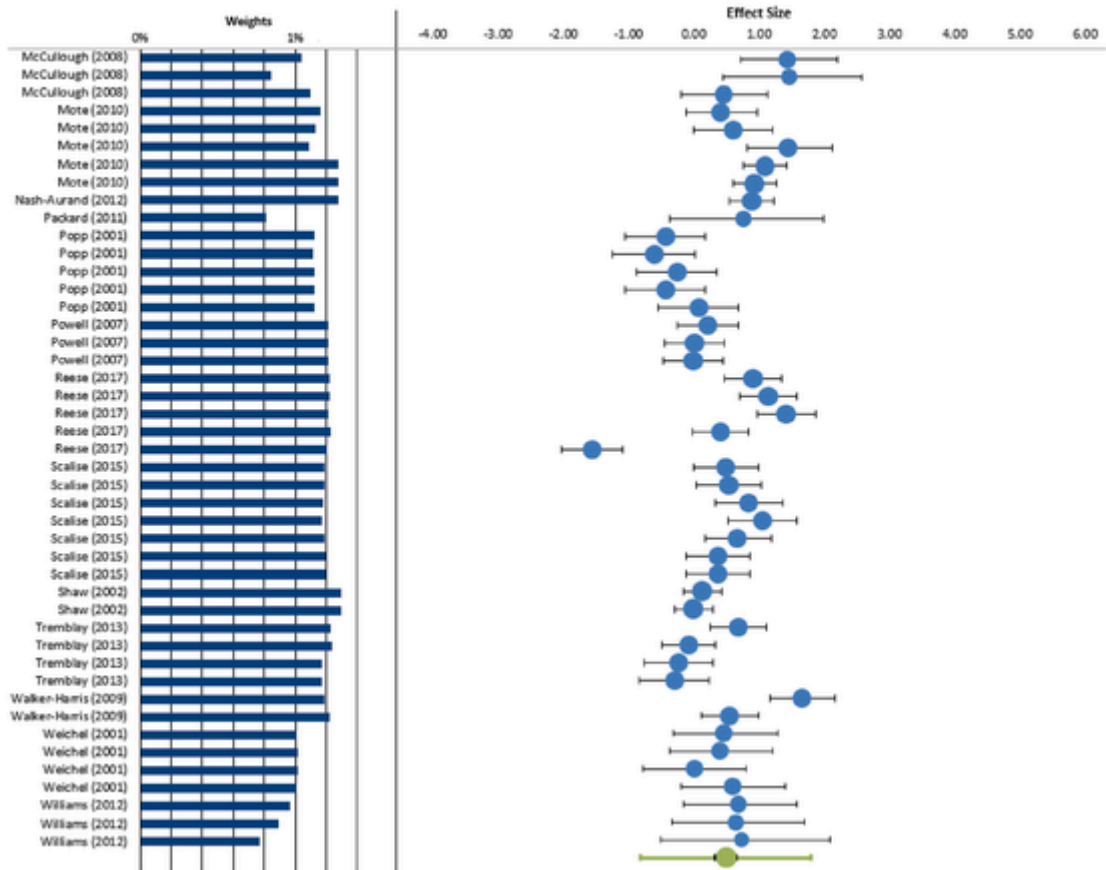


Fig. 2b. Forest plot (part II).

Lastly, we investigated whether effects of co-teaching differed between unpublished and published studies. The subgroup analysis did not indicate statistically significant differences ($p = 0.059$). However, the effect sizes reported in the articles were lower ($g = 0.22$; 95% C.I. [-0.01, 0.46]; weight = 42.61%; $I^2 = 47.49%$; $\tau^2 = 0.07$) than those reported in the dissertations ($g = 0.49$; 95% C.I. [0.37, 0.62]; weight = 57.39%; $I^2 = 84.03%$; $\tau^2 = 0.30$).

3.3. Moderator analyses

To further explore the role of contextual factors in the academic achievement of students in co-taught compared to special education classes, we conducted moderator analyses, which allowed us to assess the impact of specific factors on study effect size. The following eight moderators were assessed in our analyses: (a) 'student grade level' (elementary, middle/high school), (b) 'content area tested in the exam' (language arts/literacy, math), (c) 'exam type' (norm-referenced, criterion-referenced) (d) 'exam classification' (non-state, state) (e) 'sampling research methodology' (convenience sampling, randomization/other), (f) 'publication year' (continuous variable), (g) 'quality of the study' (continuous variable), and (h) 'publication status' (unpublished, published).

We performed moderator analyses only if at least ten observations (for students' achievement characteristics) or at least ten studies (for study characteristics) were included in each subgroup (Deeks et al., 2019). For study characteristics, we synthesized any dependent effects (Hedges, 2019), to analyze the independent effects across studies. Our analysis revealed only one statistically significant moderator: the 'student grade level.' In terms of grade level, co-teaching was significantly more beneficial for middle school and high school students ($B = 0.283$, $SE = 0.14$, $df = 87$, $p = 0.049$, C.I. lower limit = 0.00; C.I. upper limit = 0.57) rather than for elementary school students.

3.4. Publication bias

Although our comprehensive search for gray literature found that dissertations comprised the majority of eligible studies, and peer-reviewed journal articles were not overrepresented, we still accounted for the potential of other sources of bias, inclusive of language (Borenstein et al., 2011). To assess possible impact of publication bias on our results, we performed a funnel plot analysis

(Pigott & Polanin, 2020). As shown in Fig. 3, the funnel plot did not reveal an extreme occurrence of asymmetry regarding the distribution of larger-sample and smaller-sample studies in relation to the effect sizes. Our findings from the Egger regression (Egger et al., 1997) indicated slight possibility of publication bias ($p = 0.049$). To further assess this, we employed the Begg and Mazumdar's test (Begg & Mazumdar, 1994), which did not render an indication of statistical significance ($p = 0.752$). Additionally, Rosenthal's 'fail-safe N' suggested the need for at least 8630 additional studies to be included in the analysis, before the research findings would be significantly impacted by publication bias (Borenstein et al., 2011). Moreover, no differences were detected between the imputed and the observed meta-analytic effect sizes. Specifically, the trim and fill method did not lead to the imputation of any data points in the funnel plot, implying that if the results were fully symmetrically distributed around the mean, the observed effect would still not change significantly (Duval & Tweedie, 2000). Finally, the sensitivity of the estimated effect was assessed (Rudolph et al., 2020). We first removed gradually, one by one, the five achievement outcomes with the strongest positive effects, which ranged from $g = 1.41$ to $g = 1.65$. We also later removed one achievement outcome with the most extreme negative effect, which was $g = -1.56$. None of these analyses concerning the sensitivity of our results led to a significant change in the size or direction of the effect, nor was symmetry of the funnel plot severely undermined. We also considered to remove any achievement outcomes that presented high weights. However, screening our data, we verified that all included achievement points had small weights, lower than 1.34%. Overall, we concluded that our meta-analytic findings were not significantly impacted by severe risks of publication bias.

4. Discussion

This meta-analysis focused on academic achievement for SWD who received instruction in co-taught versus special education settings. The latter consisted of resource, self-contained, or a combination of resource and self-contained classes in the USA and Belgium. Importantly, we acknowledge that special education placement terminology varies across international contexts. To the extent that international contexts are similar to those described in this meta-analysis, a number of our findings can be useful on a broader scope. Twenty-six studies with a total 3,714 students, comprised of 1,975 students in co-taught settings and 1,739 students in special education settings, met eligibility criteria. In line with our first research question, we concluded achievement for SWD differs between co-taught and special education settings. Per our second research question, we identified a number of factors that differentiate SWD's achievement between co-taught and special education settings, with student grade level being a significant moderator.

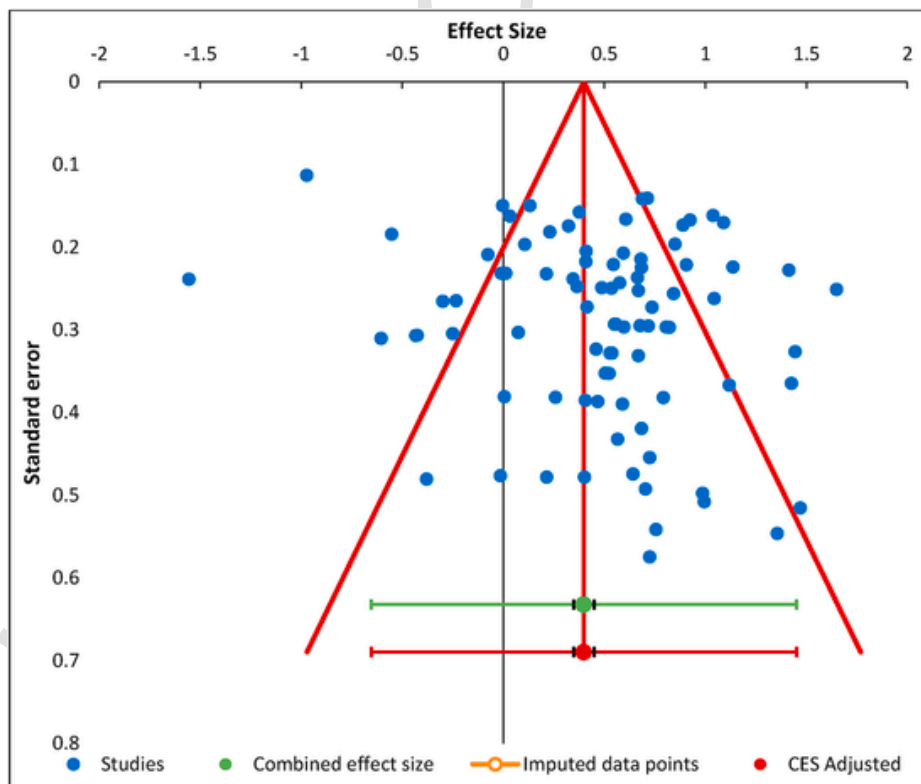


Fig. 3. Publication bias funnel plot.

4.1. Co-taught versus special education settings

The effect size for co-taught SWD compared to SWD from special education settings was medium positive ($g = 0.47$). Similarly, [Khoury \(2014\)](#) found co-taught SWD performed better compared to one type of special education setting, the resource room ($g = 0.45$). He defined resource rooms as special education settings where students received “specially designed instruction,” which he noted was at a “slower pace” than in general education settings (p. 17). Although our analysis also included SWD in self-contained settings, findings from both studies support that SWD experience notable academic benefits in co-taught versus special education settings.

Similar to our finding, [Losinski et al. \(2019\)](#) have reported an omnibus effect size ($d = 0.260$) implying that co-teaching has a positive impact for SWD. For the nine studies examined in the Losinski et al. meta-analysis, which focused only on secondary grades, results for co-teaching compared to resource rooms yielded a smaller, albeit still positive, effect. We suggest that the additional studies we acquired via the comprehensive search of the gray literature and the inclusion of all grades contributed to our larger effect size. Because [Murawski and Swanson's \(2001\)](#) meta-analysis with six studies combined data for students with and without disabilities, comparison with our study was prohibitive.

4.1.1. Special education setting and group assignment

Examination of the two different types of special education settings revealed medium effect sizes for academic achievement of co-taught SWD compared to peers in resource ($g = 0.41$), self-contained ($g = 0.51$), or in combinations of resource and self-contained ($g = 0.50$) settings. No significance for student achievement in co-taught settings was observed ($p = 0.693$), indicating that students achieved similarly regardless of type of special education setting.

In our analysis, we also sought to explore whether SWD with lower achievement assigned to special education settings would perform as well as SWD with low achievement in co-taught settings. Because how students are assigned to the different settings may depend on students' characteristics, it was important to discern whether convenience or random sampling (or comparable, such as matched pairs) methodology was used across studies.

For group assignment to either co-teaching or special education settings, there was a significant difference ($p = 0.022$) in effect sizes for studies that used convenience sampling versus randomization or similar equivalency techniques. Studies using convenience sampling had a larger effect size ($g = 0.51$) than studies that used some type of group equivalency, such as randomization ($g = 0.24$). Two studies were explicit that the student groups were not equivalent. [Beachum \(2016\)](#) noted that SWD in resource rooms were more severely impacted by their disabilities; school personnel placed SWD more likely to succeed in co-taught classes. Relatedly, [McCullough \(2008\)](#) noted SWD's placement was dependent on whether they were graduating with a regular diploma, specifying co-taught settings, or with a certificate of attendance, specifying resource settings. Studies that used intact groups of SWD in either setting sometimes noted students' current placements prohibited randomly assigning students (e.g., [Holmes, 2018](#); [Nash-Aurand, 2013](#); [Packard et al., 2011](#)) or that students' Individualized Education Programs prescribed the placement (e.g., [Gerlach, 2017](#); [Laffitte, 2012](#); [Mote, 2010](#); [Williams, 2012](#)). Other studies randomly assigned schools to treatment (e.g., [Bottge et al., 2018](#)) or established comparability based on several variables, such as gender and ethnicity ([Beam, 2005](#)). When randomization or similar techniques for group equivalency were used (e.g., [Powell, 2007](#)), students in the co-taught group benefitted, thus lending more confidence in our finding that SWD in co-taught fared better academically than their peers in special education settings. Hence, we maintain that randomization or similar methods increase the probability of identifying more reliable differences with regard to co-teaching efficacy, and, thus, these methods should be prioritized when designing studies.

4.1.2. Academic achievement measures

We performed two examinations regarding measures. The first compared results between the two groups when norm-referenced versus criterion-referenced tests were used, and the second focused on comparisons for students' performance in state versus non-state tests. For the first examination, SWD in co-taught benefitted statistically significantly more ($p = 0.049$) when they took criterion-referenced tests ($g = 0.55$) versus norm-referenced tests ($g = 0.32$). The former are tightly aligned with each subject per grade-level curriculum, and subsequently are more likely to match grade-level instruction, whereas the latter are based on more global items. For example, [Bottge et al. \(2015\)](#) found differences in the academic achievement of SWD and SWOD who were in treatment co-taught classes compared to those in control co-taught classes, depending on whether scores were derived from norm- versus criterion-referenced tests. Although in both types of tests all students from treatment co-taught classes performed better, similarly to our finding, the authors identified statistically significant better performance among students who took a criterion-referenced test.

Results supporting students' achievement from criterion- versus norm-referenced tests vary across individual studies and meta-analyses. For example, in a meta-analysis by [Graham et al. \(2018\)](#), no statistical difference was detected for norm- versus criterion-referenced tests. Conversely, in a meta-analysis by [Stockard et al. \(2018\)](#), norm-referenced tests' effect sizes were significantly larger in some content areas, but smaller in others, when compared to criterion-referenced tests. Finally, some meta-analyses have excluded studies that had not used norm-referenced measures (e.g., [Austin et al., 2019](#); [Scammacca et al., 2015](#)), precluding more comprehensive comparisons. We conclude for our study that the benefits of co-teaching are significantly larger when SWD take criterion-referenced tests. We also submit that criterion-referenced tests' content, closely aligned with grade-level curriculum, is emphasized more in co-taught versus special education settings.

For the second analysis, results indicated similar effect sizes ($p = 0.919$) among co-taught SWD who completed state tests ($g = 0.48$) and non-state tests ($g = 0.43$). State tests primarily encompass the curriculum topics within specific content areas (e.g.,

Biology), whereas the scope of non-state tests tends to be broader. For both cases, our data indicate that co-taught SWD benefit comparably for both state and non-state tests.

4.2. Grade levels, content areas, and school location

4.2.1. Grade levels

Sub-group comparisons among grade levels for SWD revealed that the effect size of co-teaching was larger ($p = 0.032$) among co-taught students in middle school ($g = 0.56$) and high school ($g = 0.52$), whereas the effect was smaller for elementary grades ($g = 0.25$). The moderator analysis rendered statistical significance ($p = 0.049$) regarding the benefits of co-teaching middle and high school SWD compared to elementary school SWD. Hence, we conclude that students in higher grade levels benefited significantly more from being in co-taught settings. Research conclusions from prior meta-analyses have been mixed. In [Khoury's \(2014\)](#) meta-analysis, there was not an observed significant effect regarding grade levels. Other meta-analyses about academic achievement did not provide comparisons between grade levels. [Losinski et al. \(2019\)](#) limited their inclusion criteria to secondary education settings, and [Murawski and Swanson's \(2001\)](#) analysis of six studies combined data, not allowing for further comparisons. In a more narrow-scope meta-analysis about the impact of reading interventions on reading self-efficacy, [Unrau et al. \(2018\)](#) reported that, based on seven studies, elementary grade interventions were more efficient than those in higher grade levels. Yet, [Szumski et al. \(2017\)](#) have reported that inclusive education among SWOD has been better for secondary grade levels ($d = 0.20$) rather than for elementary grade levels ($d = 0.11$), albeit not at a statistically significant level ($p = 0.14$). We reflect that our meta-analysis is the first one to include all grade levels with a broad array of content areas, thus allowing for a more meaningful breakdown and more robust comparison of data.

We posit that students' developmental levels, teachers' expectations of learners, and how curriculum is delivered vary significantly from elementary to middle to high school ([Gindi, 2020](#)). Because elementary SWD undergo eligibility and placement processes, which may attenuate performance differences between co-taught and non-co-taught students, we further note that these processes may result in diminishing the size of co-teaching effects for students from lower grade levels. As [Schulte et al. \(2016\)](#) noted, SWD's reading growth made "larger gains in the early grades which decelerated as students transitioned across grades" (p. 937), possibly increasing performance discrepancies between co-taught and non-co-taught SWD in secondary grades. Moreover, because elementary students' tenure in special education is at the startup stage, their performances may be closer to their peers than when SWD are in middle and high school grades. Additionally, many elementary SWD enter then exit special education services ([Woods, 2020](#)), particularly during fourth grade ([Nese et al., 2017](#)), thus indicating less of a gap between their achievement and that of their peers in other co-taught elementary classes.

An additional reason regarding discrepancies between middle schools and elementary schools relates to the fact that middle schools' design and philosophical approaches, including their interdisciplinary instruction ([Moser et al., 2019](#)), collaboration among teachers ([Schaefer et al., 2016](#)), and personalized learning for all students ([Kiefer & Ellerbrock, 2010](#); [Nagle & Taylor, 2017](#)), are well matched to co-teaching. For middle school students in their early adolescent development stage ([Booker, 2018](#)), the co-taught environment promotes efficient learning of the general education curriculum. Moreover, by high school, SWD have determined whether to seek a certificate of attendance or a regular high school diploma. Seeking a diploma means SWD need to pass state assessments for specific curricula (e.g., English; Algebra), which is the foremost emphasis in co-taught classes. Conversely, SWD in special education settings explore career opportunities, which may not require a high school diploma. Altogether, SWD in middle and high school co-taught settings are distinctly different than their peers in special education settings. Hence, these discrepancies between students in special education and co-taught settings make the magnitude of the positive benefits derived from secondary-school co-teaching significantly larger.

4.2.2. Content areas

With regard to subjects, the sub-group analysis ($p = 0.052$) showed that co-teaching presented medium effect sizes for language arts/literacy ($g = 0.60$) and mathematics ($g = 0.42$), whereas the positive effect was weak ($g = 0.08$) for other subjects (e.g., science, social studies). For [Khoury \(2014\)](#), who configured subjects differently, there were significant effects for co-taught SWD in verbal/language arts ($g = 0.330$) and math/science ($g = 0.226$), albeit lower than those we found. Other studies support that students in general education, including co-taught settings, outperform their peers in special education classes for specific content areas. For example, [Barrett et al. \(2020\)](#) found that the more time SWD spent in general education, the higher their performance on English language arts and math assessments was. Similarly, [Cole et al. \(2020\)](#) found SWD who received 80% or more of their instruction in general education performed significantly better on reading and math state tests.

4.2.3. School location

Co-taught SWD who attended rural, urban, or suburban schools differed in academic achievement from their special education peers, albeit the difference was not statistically significant ($p = 0.059$). Co-teaching's benefits in the academic achievement for co-taught SWD in suburban schools was lower ($g = 0.28$) compared to the higher effect sizes for SWD in urban ($g = 0.57$) and rural schools ($g = 0.64$). Additionally, for studies in which students were from the combination of urban, suburban, or rural schools, the effect size was even weaker ($g = 0.12$) compared to the other effect sizes. We observed that, in suburban schools, co-teaching's effect was smaller, whereas differences between co-taught and special education SWD's achievement were greater in rural and urban settings. Rural students benefitted from co-teaching the most compared to students from other school settings. Given that certain studies

only provided a classification that combined ‘urban, suburban, or rural schools,’ we posit that research which disaggregates findings by explicitly noting the school location would allow for informative comparisons and for the extraction of more robust meta-analytic conclusions.

Sindelar et al. (2018) noted that rural and urban districts have significant special education teacher shortages. Exacerbating those shortages in rural areas is that higher percentages of students are eligible for special education services (Showalter et al., 2017). All told, there are more SWD in rural special education settings who are poor, transient, and lack sufficient certified special educators, subsequently impacting students' academic achievement. Similar factors affect SWD in urban settings, with higher poverty and family mobility rates than in suburban settings (Logan & Burdick-Will, 2017; Rude & Miller, 2018). Fewer students live in poverty in suburban settings, and ‘patterns of test performance favor students in suburban schools, with reading and test scores lowest in urban schools and nearly as low in rural schools’ (Logan & Burdick-Will, 2017, p. 207). This pattern was not mirrored in the findings of our meta-analysis, with the effect of co-teaching in urban and rural schools being higher than that in suburban schools. Very importantly, our results indicate that, in spite of the hardships associated with rural areas, SWD from these areas benefited from co-teaching the most. We maintain that how co-teaching is operationalized between and among these locations be richly described in future studies, so that researchers can discern pedagogies and instructional arrangements that are most beneficial for SWD from diverse settings.

4.3. Publication and study quality factors

4.3.1. Publication

Two aspects of studies' publication were examined. First, earlier and more recent studies were compared and second, whether the publication was in a peer-reviewed journal or from gray literature (i.e., dissertation) was analyzed. For currency of the study, division occurred by decades (1999–2009; 2010 to 2019). Effect sizes were similar for each decade, with more current studies slightly higher ($g = 0.49$) than the earlier studies ($g = 0.43$). None of the previous meta-analyses on co-teaching (e.g., Khoury, 2014) analyzed studies by publication dates.

For publication type (peer-reviewed journal or gray literature), statistically significant differences ($p = 0.059$) were not obtained. Even so, effect sizes for studies published in peer-reviewed journals were lower ($g = 0.22$) than those for dissertations ($g = 0.49$). Correspondingly, Khoury's (2014) meta-analysis found significance favoring dissertations ($g = 0.439$), with negative effect size for peer-reviewed journals ($g = -0.343$). In contrast, in the meta-analysis conducted by Losinski et al. (2019), there were no meaningful differences. In other meta-analyses, Graham et al. (2018) found no statistical difference for published versus unpublished studies. Interestingly, there is considerable research noting that findings from published studies result in both positive and larger effect sizes when compared to gray literature sources (e.g., Cheung & Slavin, 2016; Conn et al., 2003).

4.3.2. Study quality

Co-teaching's effect sizes for studies with lower quality ($g = 0.52$) was comparable to that of studies with higher quality ($g = 0.45$), with no correlation between quality scores and effect sizes. This lack of correlation matches findings from other researchers (Graham et al., 2018; Kim et al., 2019). Moreover, some researchers advise less reliance on quality scores that may represent authors' subjective ratings (Nguyen et al., 2019; Sweigart et al., 2016), and suggest that some quality items may be underreported (Marusic et al., 2020). Nonetheless, multiple scholars urge researchers to objectively acquire and subsequently report study quality ratings (Appelbaum et al., 2018; Cook et al., 2014; Gersten et al., 2005; Talbott et al., 2017; Valentine & Cooper, 2008), even though multiple meta-analyses omit them (Namkung et al., 2019; Richter et al., 2016; Spencer & Wagner, 2018).

4.3.2.1. Participant descriptions. Although more descriptive information was provided for SWD than for teachers (e.g., 42% ethnicity and gender reported for students; no ethnicity and 9% gender reported for teachers), demographic content was incomplete across all studies. Incomplete information on teacher participants is reported by others, such as by Didion et al. (2020), who note in their meta-analysis on teacher professional development that few studies report teacher experience and certification. In particular for co-teachers, because preparation to co-teach is crucial for effective co-teaching (Embury & Dinnesen, 2013), only a few studies (19%) reported training. Pancsofar & Petroff (2013) describe multiple benefits for co-teachers, and subsequently their students, due to frequent training opportunities. Conversely, multiple studies note co-teachers who lack sufficient preparation have lower self-efficacy, which impacts their effectiveness (Fennick & Liddy, 2001; Friend et al., 2010).

4.3.2.2. Co-teaching descriptions. All studies noted co-teaching consisted of general and special educators teaming together in one classroom setting, but beyond that, details were not reported in many studies (see Table 1). Sorely underreported is how co-teaching was operationalized as an independent variable; only 19% of studies described co-teachers' actions. Interestingly, other meta-analyses, such as on reading interventions for at-risk students in primary grades, a well-researched topic, also reported omission of sufficient independent variable descriptions in studies (Gersten et al., 2020): ‘‘Only a few studies ... provided a rich description on the nature of instruction in the intervention. Most articles did not provide sufficient detail on how reading is taught; ... [leading to] coding or classifying these areas was at times guesswork’’ (p. 421). Similarly, instruction in special education settings was not described beyond noting special educators as the teachers.

Co-teaching models were identified in only 23% of studies, and the majority (92%) of studies entirely omitted fidelity. Similar omissions are reported by all researchers on meta-analyses about co-teaching (Khoury, 2014; Losinski et al., 2019; Murawski & Swanson, 2001).

4.3.2.3. Effective co-teaching. Because most studies omitted description and corresponding fidelity measures, we must be cautious about the interpretation of our findings, even though our data clearly support higher achievement for co-taught SWD. The majority of the authors of the examined studies did not provide sufficient information about how co-teaching occurred. Previous studies that include descriptions about how co-teaching was operationalized have substantiated effective (e.g., Kim et al., 2006; King-Sears et al., 2014; Shaffer & Thomas-Brown, 2015) and ineffective co-teaching (Keefe & Moore, 2004; Magiera & Zigmond, 2005; Wexler et al., 2018). Other co-teaching studies have raised concerns regarding parity between the co-teachers, such as when general educators frequently lead most instruction (Cook et al., 2017; Pancsofar & Petroff, 2013; Scruggs et al., 2007), which has also been noticed by SWD (Bessette, 2008; Preston-Smith et al., 2020).

Similarly to our observations, Murawski and Swanson (2001) acquired information about effective co-teaching in their six studies, such as content about exhibiting parity, volunteering to co-teach, sharing resources and responsibilities, and using different co-teaching models. Although some of this content was interspersed across studies, none included all elements. Losinski et al. (2019) and Khoury (2014) also noted the omission of sufficient descriptors regarding co-teaching processes. Several authors have previously supported that effective co-teaching must be preceded by sufficient co-planning, which Stefanidis et al. (2019) found co-teachers profited from, a finding that is endorsed by multiple researchers (e.g., Dusty & Schneider, 2012; Strogilos et al., 2016).

Consequently, we are emphatic that our meta-analysis results supporting academic achievement for co-taught SWD are not accompanied by information indicating the components of instruction in co-taught settings. Similarly, the elements of instruction in the special education settings are not described, nor discussed. Absent clarity on and evidence about instruction being provided to SWD, we cannot claim that high-quality instruction was or was not occurring for SWD in either setting. Multiple studies, including previous meta-analyses, acknowledge that lack of information about the instruction in co-taught or special education settings preclude identification of research-based practices used with fidelity (Khoury, 2014; Losinski et al., 2019; Murawski & Swanson, 2001). Hence, similarly to Losinski et al. (2019), we call for high-quality studies that contrast “state-of-the-art co-teaching versus state-of-the-art instruction” in special education settings (p. 158). We concur with Zigmond et al.’s (2013) perspective specifying co-taught settings as places to deliver services to SWD and SWOD, rather than co-teaching as a stringent set of interventions. Moreover, we further maintain that such services should be comprised of research-based pedagogies implemented by two fully-certified teachers, one in general education and the other one in special education. As such, services comprised of these pedagogies and varied instructional arrangements (i.e., co-teaching models) do lend themselves to co-teaching practices implemented with fidelity. While co-teaching is not implemented identically in different classes (e.g., English versus Science), there are research-based practices that should be evident within and across content areas (Bottge et al., 2015; Freeman et al., 2021; Lory et al., 2020; Therrien et al., 2011). To that end, we posit that, while fidelity measures may vary, effective co-teaching should be performed as intended (i.e., with fidelity), incorporating detailed ad hoc planning, followed by subsequent implementation and post hoc monitoring processes.

4.4. Limitations, future research, and conclusions

As with any meta-analysis, unpublished studies outside the broad international scope with which we searched may not have been acquired. Although we were comprehensive in our search, we are reliant on researchers’ awareness of and responsiveness to calls for unpublished studies. Future researchers should continue to include thorough searches for international gray literature. Omitting gray literature impacts inclusion of the universe of studies. As occurred in our search, there were data available, but many were not in peer-reviewed journals, and some manuscripts were acquired from work authors had within the review process. In relation to this, we urge that our findings be considered in view of the notion that special education placements vary within and across international contexts, and we invite future researchers to integrate this consideration in their work regarding student academic outcomes. We explicitly acknowledge that almost all of our meta-analyzed data derived from the USA. Although our findings may present value for scholars and practitioners in broader international contexts, nonetheless, we urge researchers from other countries to assess generalizability to their settings. Another limitation is that we constrained our comparison groups to SWD in two specific settings: co-taught and special education resource and self-contained. We did not seek to compare academic achievement of SWD versus SWOD because we anticipated students’ characteristics would be too variant. For example, Gilmour et al.’s (2019) meta-analysis that compared reading achievement of SWD to that of SWOD yielded variability across grade levels, measurements, and disability categories. Further, the available information regarding teachers’ demographics and other relevant information, such as certifications, professional development training, and co-teaching planning, were limited. We encourage all researchers who conduct co-teaching studies to report more comprehensive information both about participants and followed co-teaching instructional processes. Consistent with our examination of all grade levels in our analysis, we recommend future researchers to continue the same disaggregation to disentangle student achievement differences between and among grade levels. Additionally, provision of detailed information regarding the disability category as well as relevant characteristics (e.g., extent of support needs) can allow for generalizability of specific research findings and for future comparisons by other meta-analytic studies.

The differential effect of co-teaching regarding the academic achievement of students as this is tested across content areas (i.e., ‘language arts/literacy,’ ‘math,’ and ‘other’) needs to be researched by future scholars to further identify variations in co-teaching efficacy depending on subjects. Relatedly, a recommendation for future researchers is to collect and meta-analyze data that would further elucidate the potential efficacy of co-teaching training across different content areas. For example, comparing SWD’s achieve-

ment before and after training interventions targeted toward their co-teachers would expand comparisons, potentially revealing the benefits of teachers' training, with the goal to increase preparation, implementation, and assessment of co-teaching practices. Assessing the effects of training for co-teachers across different content areas has the potential to enrich knowledge regarding co-teaching efficacy.

How well SWD fare in co-taught compared to special education settings from an academic achievement perspective is only one element representing the progress, well-being, and social world of SWD. We did not seek measures on affective factors in this meta-analysis, which is a recommendation for future researchers, so that they more fully discern other important aspects indicative of SWD's performance across multiple domains of school life, some of which can be predictors for their future. For example, Sanders et al. (2018) noted that a positive school climate could impact the achievement for SWD at all grade levels and in varied school types (e.g., urban). Similarly, Garrote et al. (2017) noted the importance of social participation of SWD in general education settings, such as friendships and social acceptance. Consequently, future meta-analyses should assess students' emotional and social benefits derived from their participation in the co-taught classroom.

Where students with disabilities are educated and how well they perform academically is an international issue (Fuchs et al., 2015; Lehane & Senior, 2020; Lemons et al., 2018; Pulkkinen et al., 2020; Weiss & Glaser, 2021). Data supporting that academic achievement for co-taught SWD is higher than that of SWD in special education settings only scratches the surface of information needed to substantiate that SWD are receiving appropriate educational services in the least restrictive environment, whatever that setting may be. Knowing where SWD received instruction does not inform how students received instruction. To be more completely informative, the "how" of instruction must be coupled with the "place" of instruction. Additionally, examinations that focus on global well-being aspects of SWD relative to their placement must be factored into a more complete picture of benefits they accrue. Ultimately, SWD need to be acquiring effective individualized instruction during their school years that promotes their personal and professional productivity as future citizens, aspects that should be more sufficiently elucidated by current studies in the field.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Uncited references

;

CRediT authorship contribution statement

Margaret E. King-Sears: Conceptualization, Data curation, Methodology, Writing – original draft, Writing – review & editing, Project administration. **Abraham Stefanidis:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Sheri Berkeley:** Conceptualization, Methodology, Data curation, Writing – review & editing. **Vasilis Strogilos:** Data curation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

None.

Acknowledgement

The authors gratefully acknowledge the following students who assisted with tasks during this research: Melissa DeLury, Alicia Ellis, Dung Ngo, Kylee Vasil, and Anastasia Voulagka.

References

- = study included in the meta-analysis
- * Adams, S. S. (2014). *Coteaching in secondary special and general education classrooms and student mathematics achievement* (Publication No. 3643563) Doctoral dissertation. Walden University]. ProQuest Dissertations & Theses Global. <https://www.proquest.com/docview/1622473702?accountid=14541>.
- * Andrews-Tobo, R. (2009). *Coteaching in the urban middle school classrooms: Impact for students with disabilities in reading, math, and English/language arts classrooms* (Publication No. 3344906) [Doctoral dissertation. Capella University]. ProQuest Dissertations & Theses Global. <https://www.proquest.com/docview/305163048/se-2?accountid=14541>.
- 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>.
- Austin, C. R., Wanzek, J., Scammacca, N. K., Vaughn, S., Gesel, S. A., Donegan, R. E., & Engelmann, M. L. (2019). *The relationship between study quality and the effects of supplemental reading interventions: A meta-analysis*. *Exceptional Children*, 48(5), 347–366. <https://doi.org/10.1177/0014402918796164>.
- Barrett, B. A., Stevenson, N. A., & Burns, M. K. (2020). *Relationship between disability category, time spent in general education and academic achievement*. *Educational Studies*, 46(4), 497–512. <https://doi.org/10.1080/03055698.2019.1614433>.
- * Beachum, L. M. (2016). *Co-teaching as an effective instructional delivery model in secondary schools* (Publication No. 10246557) [Doctoral dissertation. Wingate University]. ProQuest Dissertations & Theses Global. <https://proquest.com/docview/1882305464?accountid=14541>.

- * Beam, A. P. (2005). *The analysis of inclusion versus pullout at the elementary level as determined by selected variables* (Publication No. 3161573) [Doctoral dissertation. The George Washington University]. ProQuest Dissertations & Theses Global. <https://proquest.com/docview/304999265?accountid=14541>.
- Begg, C. B., & Mazumdar, M. (1994). *Operating characteristics of a rank correlation test for publication bias*. *Biometrics*, 50(4), 1088–1101. <https://doi.org/10.2307/2533446>.
- Bessette, H. J. (2008). *Using students' drawings to elicit general and special educators' perceptions of co-teaching*. *Teaching and Teacher Education*, 24(5), 1376–1396. <https://doi.org/10.1016/j.tate.2007.06.007>.
- Booker, K. (2018). *The high tide raises all ships: Middle grades teachers' perspectives on school belonging in early adolescence*. *RMLE Online*, 41(8), 1–15. <https://doi.org/10.1080/19404476.2018.1505402>.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons.
- * Bottge, B. A., Cohen, A. S., & Choi, H. J. (2018). *Comparisons of mathematics intervention effects in resource and inclusive classrooms*. *Exceptional Children*, 84(2), 197–212. <https://doi.org/10.1177/0014402917736854>.
- Bottge, B. A., Toland, M. D., Gassaway, L., Butler, M., Choo, S., Griffen, A. K., & Ma, X. (2015). *Impact of enhanced anchored instruction in inclusive math classrooms*. *Exceptional Children*, 81(2), 158–175. <https://doi.org/10.1177/0014402914551742>.
- Brown, M. J., & Sutton, A. J. (2010). *Quality control in systematic reviews and meta-analysis*. *European Journal of Vascular and Endovascular Surgery*, 40(5), 669–677. <https://doi.org/10.1016/j.ejvs.2010.07.011>.
- Brusca-Vega, R., Brown, K., & Yasutake, D. (2011). *Science achievement of students in co-taught, inquiry-based classrooms*. *Learning Disabilities: A Multidisciplinary Journal*, 17(1), 23–31.
- Burke, K. M., Raley, S. K., Shogren, K. A., Hagiwara, M., Mumbardo-Adam, C., Uyanik, H., & Behrens, S. (2020). *A meta-analysis of interventions to promote self-determination for students with disabilities*. *Remedial and Special Education*, 41(3), 176–188. <https://doi.org/10.1177/0741932518802274>.
- * Castro, V. E. (2007). *The effect of co-teaching on academic achievement of K–2 students with and without disabilities in inclusive and non-inclusive classrooms* (Publication No. 3255067) [Doctoral dissertation. Fordham University]. ProQuest Dissertations & Theses Global. <https://proquest.com/docview/304883065?accountid=14541>.
- Cheung, A., & Slavin, R. (2016). *How methodological features affect effect sizes in education*. *Educational Researcher*, 45(5), 283–292. <https://doi.org/10.3102/0013189X16656615>.
- * Clements, T. P. (2012). *How does inclusion affect African American middle school special education student performance? A comparison of disability categories* (publication No. 3514804) [Doctoral dissertation. Union University]. ProQuest Dissertations & Theses Global. <https://proquest.com/docview/1021723748?accountid=14541>.
- Cole, S. M., Murphy, H. R., Frisby, M. B., Grossi, T. A., & Bolte, H. R. (2020). *The relationship of special education placement and student academic outcomes*. *The Journal of Special Education*, 54(4), 217–227. <https://doi.org/10.1177/0022466920925033>.
- Cole, C. M., Waldron, N., & Majd, M. (2004). *Academic progress of students across inclusive and traditional settings*. *Mental Retardation*, 42(2), 136–144. [https://doi.org/10.1352/0047-6765\(2004\)42<136:APOSAI>2.0.CO;2](https://doi.org/10.1352/0047-6765(2004)42<136:APOSAI>2.0.CO;2).
- Conn, V. S., Valentine, J. C., Cooper, H. M., & Rantz, M. J. (2003). *Grey literature in meta-analyses*. *Nursing Research*, 52(4), 256–261. <https://doi.org/10.1097/00006199-200307000-00008>.
- Cook, B. G., Buysse, V., Klingner, J., Landrum, T. J., McWilliam, R. A., Tankersley, M., & Test, D. W. (2014). *CEC's standards for classifying the evidence base of practices in special education*. *Remedial and Special Education*, 36(4), 220–234. <https://doi.org/10.1177/0741932514557271>.
- Cook, S. C., McDuffie-Landrum, K. A., Oshita, L., & Cook, B. G. (2017). *Co-teaching for students with disabilities: A critical and updated analysis of the empirical literature*. In J. M., Kauffman, D. P., Hallahan, & P. C., Pullen (Eds.), *Handbook of special education* (pp. 233–248). Routledge.
- Deeks, J. J., Higgins, J. P., Altman, D. G., & Cochrane Statistical Methods Group (2019). *Analysing data and undertaking meta-analyses*. In P. T., Higgins, J., Thomas, J., Chandler, M., Cumpston, T., Li, M. J., Page, & V. A., Welch (Eds.), *Cochrane handbook for systematic reviews of interventions* (2nd ed., pp. 241–284). The Cochrane Collaboration. <https://doi.org/10.1002/9781119536604.ch10>.
- Department of Education Area guidelines for SEND and alternative provision including special schools, alternative provision, specially resourced provision and units Accessed 11th May 2021 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/905693/BB104.pdf 2015
- Department for Education Special educational needs in England: January 2020 Accessed 11th May 2021 <https://www.gov.uk/government/statistics/special-educational-needs-in-england-january-2020> 2020
- Didion, L., Toste, J. R., & Filderman, M. J. (2020). *Teacher professional development and student reading achievement: A meta-analytic review of the effects*. *Journal of Research on Educational Effectiveness*, 13(1), 29–66. <https://doi.org/10.1080/19345747.2019.1670884>.
- * Doran, J. B., Jr. (2008). *Comparing two methods for instructing students in special education: Coteaching and small group instruction* (Publication No. 3303503) Doctoral dissertation. Walden University]. ProQuest Dissertations & Theses Global. <https://proquest.com/dissertations-theses/comparing-two-methods-instructing-students/docview/304380841/se-2?accountid=14541>.
- Dusty, C. E., & Schneider, M. (2012). *Co-teaching in inclusive classrooms using structured collaborative planning*. *Kentucky Journal of Excellence in College Teaching & Learning*, 10, 36–52.
- Duval, S., & Tweedie, R. (2000). *Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis*. *Biometrics*, 56(2), 455–463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>.
- Embury, D. C., & Dinnesen, M. S. (2013). *Co-teaching in inclusive classrooms using structured collaborative planning*. *Kentucky Journal of Excellence in College Teaching & Learning*, 10(3), 36–52. <https://encompass.eku.edu/kjctl/vol10/iss2012/3>.
- Fennick, E., & Liddy, D. (2001). *Responsibilities and preparation for collaborative teaching: Co-teachers' perspectives*. *Teacher Education and Special Education*, 24(3), 229–240. <https://doi.org/10.1177/088840640102400307>.
- Ferguson, C. J., & Brannick, M. T. (2012). *Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses*. *Psychological Methods*, 17(1), 120–128. <https://doi.org/10.1037/a0024445>.
- * Franklin, J. C. (2015). *Co-teaching student outcomes and efficacious practices: A pragmatic mixed methods study* (publication No. 3663855) [Doctoral dissertation. Mercer University]. ProQuest Dissertations & Theses Global. <https://proquest.com/docview/1708273149?accountid=14541>.
- Freeman-Green, S., Driver, M. K., Wang, P., Kamuru, J., & Jackson, D. (2021). *Culturally sustaining practices in content area instruction for CLD students with learning disabilities*. *Learning Disabilities Research & Practice*, 36(1), 12–25. <https://doi.org/10.1111/ldrp.12240>.
- Friend, M., Cook, L., Hurlley-Chamberlain, D., & Shamberger, C. (2010). *Co-teaching: An illustration of the complexity of collaboration in special education*. *Journal of Educational and Psychological Consultation*, 20(1), 9–27. <https://doi.org/10.1080/10474410903535380>.
- Fuchs, L. S., Fuchs, D., Compton, D. L., Wehby, J., Schumacher, R. F., Gersten, R., & Jordan, N. C. (2015). *Inclusion versus specialized intervention for very-low-performing students: What does access mean in an era of academic challenge?* *Exceptional Children*, 81(2), 134–157. <https://doi.org/10.1177/0014402914551743>.
- Gage, N. A., Cook, B. G., & Reichow, B. (2017). *Publication bias in special education meta-analyses*. *Exceptional Children*, 83(4), 428–445. <https://doi.org/10.1177/0014402917691016>.
- Garrote, A., Dessesmontet, R. S., & Opitz, E. M. (2017). *Facilitating the social participation of pupils with special educational needs in mainstream schools: A review of school-based interventions*. *Educational Research Review*, 20, 12–23. <https://doi.org/10.1016/j.edurev.2016.11.001>.
- * Gerlach, S. M. (2017). *A quantitative study of co-teaching as an instrumental model to serve elementary students* [Doctoral Dissertation. St. John Fisher College]. Fisher Digital Publications. <https://fisherpub.sjfc.edu/cgi/viewcontent.cgi?article=1303&context=education>.
- 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>.
- Gersten, R., Haymond, K., Newman-Gonchar, R., Dimino, J., & Jayanthi, M. (2020). *Meta-analysis of the impact of reading interventions for students in the primary grades*. *Journal of Research on Educational Effectiveness*, 13(2), 401–427. <https://doi.org/10.1080/19345747.2019.1689591>.
- Gilmour, A. F., Fuchs, D., & Wehby, J. H. (2019). *Are students with disabilities accessing the curriculum? A meta-analysis of the reading achievement gap between students with and without disabilities*. *Exceptional Children*, 85(3), 329–346. <https://doi.org/10.1177/0014402918795830>.
- Gindi, S. (2020). *Educational placement of students with autism spectrum disorder and its relation to socioeconomic status, intelligence, and diagnosis*. *International Journal of Developmental Disabilities*, 66(3), 235–244. <https://doi.org/10.1080/20473869.2019.1569359>.

- Graham, S., Liu, X., Barlett, B., Ng, C., Harris, K. R., Aitken, A., ... Talukdar, J. (2018). *Reading for writing: A meta-analysis of the impact of reading interventions on writing. Review of Educational Research, 88*(2), 243–384.
- Hang, Q., & Rabren, K. (2009). An examination of co-teaching: Perspectives and efficacy indicators. *Remedial and Special Education, 30*(5), 259–268. <https://doi.org/10.1177/0741932508321018>.
- Harari, M. B., Parola, H. R., Hartwell, C. J., & Riegelman, A. (2020). *Literature searches in systematic reviews and meta-analyses: A review, evaluation, and recommendations. Journal of Vocational Behavior, 118*, 103377. <https://doi.org/10.1016/j.jvb.2020.103377>.
- Hedges, L. V. (2019). Stochastically dependent effect sizes. In H., Cooper, L. V., Hedges, & J. C., Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 281–297). Russell Sage Foundation.
- * Holmes, C. C. (2018). *Effect of coteaching on the achievement of students with disabilities* (Publication No. 10936503) [Doctoral dissertation. Walden, University]. ProQuest Dissertations & Theses Global. <https://proquest.com/docview/2123079412?accountid=14541>.
- van Hover, S., Hicks, D., & Sayeski, K. (2012). A case study of co-teaching in an inclusive secondary high-stakes World History I classroom. *Theory & Research in Social Education, 40*(3), 260–291. <https://doi.org/10.1080/00933104.2012.705162>.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Sage.
- Indiana State Department of Education (2021). *Special educators who are co-teachers 2010-2020. Individuals with disabilities education act, 20 U.S.C. § 1400*. 2004.
- Keefe, E. B., & Moore, V. (2004). The challenge of co-teaching in inclusive classrooms at the high school level: What the teachers told us. *American Secondary Education, 32*(3), 77–88. <https://www.jstor.org/stable/41064524>.
- Khoury, C. (2014). *The effect of co-teaching on the academic achievement outcomes of students with disabilities: A meta-analytic synthesis* [Doctoral dissertation. University of North Texas]. ProQuest Dissertations & Theses Global. <https://search-proquest-com.mutex.gmu.edu/docview/1719051534?accountid=14541>.
- Kiefer, S. M., & Ellerbrock, C. R. (2010). Understanding middle grades students' Perceptions of their peer worlds: Implications for teaming. *Middle School Journal, 42*(2), 48–54. <https://doi.org/10.1080/00940771.2010.11461756>.
- Kim, S. W., Cho, H., & Kim, L. Y. (2019). Socioeconomic status and academic outcomes in developing countries: A meta-analysis. *Review of Educational Research, 89*(6), 875–916. <https://doi.org/10.3102/0034654319877155>.
- Kim, A. H., Woodruff, A. L., Klein, C., & Vaughn, S. (2006). Facilitating co-teaching for literacy in general education classrooms through technology: Focus on students with learning disabilities. *Reading & Writing Quarterly, 22*(3), 269–291. <https://doi.org/10.1080/10573560500455729>.
- * King, A. A. (2015). *A descriptive comparative study examining school-wide co-teaching and student achievement scores* (Publication No. 3719499) [Doctoral dissertation. Capella University]. ProQuest Dissertations & Theses Global. <https://proquest.com/dissertations-theses/descriptive-comparative-study-examining-school/docview/1717109816/se-2?accountid=14541>.
- King-Sears, M. E., Brawand, A. E., Jenkins, M. C., & Preston-Smith, S. (2014). Co-teaching perspectives from secondary science co-teachers and their students with disabilities. *Journal of Science Teacher Education, 25*(6), 651–680. <https://doi.org/10.1007/s10972-014-9391-2>.
- * Laffitte, L., Jr. (2012). *A comparison of pull-out and co-teaching models on the reading performance of third through fifth grade elementary students with a diagnosed specific learning disability in reading* (Publication No. 3540689) [Doctoral dissertation. Pepperdine University]. ProQuest Dissertations & Theses Global. <https://proquest.com/dissertations-theses/comparison-pull-out-co-teaching-models-on-reading/docview/1112071602/se-2?accountid=14541>.
- Leafstedt, J. M., Richards, C., LaMonte, M., & Cassidy, D. (2007). Perspectives on co-teaching: Views from high school students with learning disabilities. *Learning Disabilities: A Multidisciplinary Journal, 14*(3), 177–184.
- Lehane, P., & Senior, J. (2020). Collaborative teaching: Exploring the impact of co-teaching practices on the numeracy attainment of pupils with and without special educational needs. *European Journal of Special Needs Education, 35*(3), 303–317. <https://doi.org/10.1080/08856257.2019.1652439>.
- Lemons, C. J., Vaughn, S., Wexler, J., Kearns, D. M., & Sinclair, A. C. (2018). Envisioning an improved continuum of special education services for students with learning disabilities: Considering intervention intensity. *Learning Disabilities Research & Practice, 33*(3), 131–143. <https://doi.org/10.1111/ldrp.12173>.
- Lipsey, M. W., & Wilson, D. B. (2000). *Practical meta-analysis*. Sage.
- Lochner, W. W., Murawski, W. W., & Daley, J. T. (2019). The effect of co-teaching on student cognitive engagement. *Theory & Practice in Rural Education, 9*(2), 6–19. <https://doi.org/10.3776/tpre.2019.v9n2p6-19>.
- Logan, J. R., & Burdick-Will, J. (2017). School segregation and disparities in urban, suburban, and rural areas. *The Annals of the American Academy of Political and Social Science, 674*(1), 199–216. <https://doi.org/10.1177/0002716217733936>.
- Lory, C., Mason, R. A., Davis, J. L., Wang, D., Kim, S. Y., Gregori, E., & David, M. (2020). A meta-analysis of challenging behavior interventions for students with developmental disabilities in inclusive school settings. *Journal of Autism and Developmental Disorders, 50*(4), 1221–1237. <https://doi.org/10.1007/s10803-019-04329-x>.
- Losinski, M., Sanders, S., Parks-Ennis, R., Wiseman, N., & Katsiyannis, A. (2019). An investigation of co-teaching to improve academic achievement of students with disabilities: A meta-analysis. *Journal of the American Academy of Special Education Professionals, 149*–170. <https://www.naset.org/index.php?id=5158>.
- Magiera, K., & Zigmond, N. (2005). Co-teaching in middle school classrooms under routine conditions: Does the instructional experience differ for students with disabilities in co-taught and solo-taught classes?. *Learning Disabilities Research & Practice, 20*(2), 79–85. <https://doi.org/10.1111/j.1540-5826.2005.00123.x>.
- Marusic, J. F., Fidahic, M., Cepeha, C. M., Farcas, L. G., Tseke, A., & Puljak, L. (2020). Methodological tools and sensitivity analysis for assessing quality or risk of bias used in systematic reviews published in the high-impact anesthesiology journals. *BMC Medical Research Methodology, 20*(121), 1–10. <https://doi.org/10.1186/s12874-020-00966-4>.
- * McCullough, J. L. (2008). *A study of special education programming and its relationship to student mathematics performance on the DSTP* (Order No. 3325487) [Doctoral dissertation. University of Delaware]. ProQuest Dissertations & Theses Global. <https://proquest.com/docview/304627371?accountid=14541>.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097. <https://pubmed.ncbi.nlm.nih.gov/19621072/>.
- Moser, K. M., Ivy, J., & Hopper, P. F. (2019). Rethinking content teaching at the middle level: An interdisciplinary approach. *Middle School Journal, 50*(2), 17–27. <https://doi.org/10.1080/00940771.2019.1576579>.
- * Mote, S. Y. (2010). *Does setting affect achievement of students with disabilities: Comparing co-teaching to resource* (Publication No. 3432877) [Doctoral dissertation. Liberty University]. ProQuest Dissertations & Theses Global. <https://proquest.com/dissertations-theses/does-setting-affect-achievement-students-with/docview/822647263/se-2?accountid=14541>.
- * Murawski, W. W. (2006). Student outcomes in co-taught secondary English classes: How can we improve?. *Reading & Writing Quarterly, 22*(3), 227–247. <https://doi.org/10.1080/10573560500455703>.
- Murawski, W. W., & Swanson, H. L. (2001). A meta-analysis of co-teaching research: Where are the data?. *Remedial and Special Education, 22*(5), 258–267. <https://doi.org/10.1177/074193250102200501>.
- Nagle, J., & Taylor, D. (2017). *Using a personal learning framework to transform middle grades teaching practice. Middle Grades Research Journal, 11*(1), 85–100.
- Namkung, J. M., Peng, P., & Lin, X. (2019). The relation between mathematics anxiety and mathematics performance among school-aged students: A meta-analysis. *Review of Educational Research, 89*(3), 459–496. <https://doi.org/10.3102/0034654319843494>.
- * Nash-Aurand, T. (2013). *A comparison of general education co-teaching versus special education resource service delivery models on math achievement of students with disabilities* (Publication No. 3588241) [Doctoral dissertation. Liberty University]. ProQuest Dissertations & Theses Global. <https://proquest.com/dissertations-theses/comparison-general-education-co-teaching-versus/docview/1443738215/se-2?accountid=14541>.
- Nese, J. F. T., Stevens, J. J., Schulte, A. C., Tindal, G., & Elliott, S. N. (2017). Modeling the time-varying nature of student exceptionalty classification on achievement growth. *The Journal of Special Education, 51*(1), 38–49. <https://doi.org/10.1177/0022466916668164>.
- Nguyen, T. D., Kramer, J. W., & Evans, B. J. (2019). The effects of grant aid on student persistence and degree attainment: A systematic review and meta-analysis of the causal evidence. *Review of Educational Research, 89*(6), 831–874. <https://doi.org/10.3102/0034654319877156>.
- * Packard, A. L., Hazelkorn, M., Harris, K. P., & McLeod, R. (2011). Academic achievement of secondary students with learning disabilities in co-taught and resource rooms. *Journal of Research in Education, 21*(2), 100–117.
- Paez, A. (2017). *Gray literature: An important resource in systematic reviews. Journal of Evidence-Based Medicine, 10*(3), 233–240. <https://doi.org/10.1111/jebm.12266>.
- Pancsofar, N., & Petroff, J. G. (2013). Professional development experiences in co-teaching: Associations with teacher confidence, interests, and attitudes. *Teacher Education and*

- Special Education*, 36(2), 83–96. <https://doi.org/10.1177/0888406412474996>.
- Pigott, T. D., & Polanin, J. R. (2020). *Methodological guidance: High-quality meta-analysis in a systematic review*. *Review of Educational Research*, 90(1), 24–46. <https://doi.org/10.3102/0034654319877153>.
- * Popp, P. A. (2001). *Standards-based assessment and program efficacy: Comparing service delivery models for students with learning disabilities and their peers without disabilities* (Publication No. 3012232) [Doctoral dissertation. The College of William and Mary]. ProQuest Dissertations & Theses Global. <https://proquest.com/docview/230758801?accountid=14541>.
- * Powell, J. E. (2007). *A comparison of learning outcomes for students with disabilities taught in three dissimilar classroom settings: Support services, team/collaborative and departmental/pullout* (Publication No. 3265518) [Doctoral dissertation. Auburn University]. ProQuest Dissertations & Theses Global. <https://proquest.com/docview/304897842?accountid=14541>.
- Preston-Smith, S., King-Sears, M. E., Evmenova, A., & Baker, P. H. (2020). *What do high school students think about co-teaching in science classrooms?. Learning Disabilities: A Multidisciplinary Journal*, 25(1), 29–42. <https://doi.org/10.18666/LDMJ-2020-V25-I1-10107>.
- Pulkkinen, J., Raikkonen, E., Jahnukainen, M., & Pirttimaa, R. (2020). *How do educational reforms change the share of students in special education? Trends in special education in Finland*. *European Educational Research Journal*, 19(4), 364–384. <https://doi.org/10.1177/1474904119892734>.
- * Reese, D. R. (2017). *Co-teaching in middle school classrooms: Quantitative comparative study of special education student assessment performance* (Publication No. 10599177) [Doctoral dissertation. Northcentral University]. Available from ProQuest Dissertations & Theses Global. 1952349105 <https://proquest.com/dissertations-theses/co-teaching-middle-school-classrooms-quantitative/docview/1952349105/se-2?accountid=14541>.
- Richter, J., Scheiter, K., & Eitel, A. (2016). *Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis*. *Educational Research Review*, 17, 19–36. <https://doi.org/10.1016/j.edurev.2015.12.003>.
- D.J. Royer K.L. Lane E.A. Common Group comparison and single-case research design quality indicator matrix using Council for Exceptional Children 2014 standards: Standards overview and walk-through guide. Unpublished tool Retrieved from <https://www.ci3t.org/practice> 2017
- Rude, H., & Miller, K. J. (2018). *Policy challenges and opportunities for rural special education*. *Rural Special Education Quarterly*, 37(1), 21–29. <https://doi.org/10.1177/8756870517748662>.
- Rudolph, C. W., Chang, C. K., Rauvola, R. S., & Zacher, H. (2020). *Meta-analysis in vocational behavior: A systematic review and recommendations for best practices*. *Journal of Vocational Behavior*, 118, 1–23. <https://doi.org/10.1016/j.jvb.2020.103397>.
- Sanders, S. M., Durbin, J. M., Anderson, B. G., Fogarty, L. M., Giraldo-Garcia, R. J., & Voight, A. (2018). *Does a rising school climate lift all boats? Differential associations of perceived climate and achievement for students with disabilities and limited English proficiency*. *School Psychology International*, 39(6), 646–662. <https://doi.org/10.1177/0143034318810319>.
- * Scalise, S. A. (2015). *An examination of the relationship between special education programs, ethnicity and socioeconomic on student growth in reading and mathematics* (Publication No. 10059807) [Doctoral dissertation. St. John's University]. ProQuest Dissertations & Theses Global. <https://proquest.com/docview/1775516929?accountid=14541>.
- Scammacca, N., Roberts, G., Vaughn, S., & Stuebing, K. (2015). *A meta-analysis of interventions for struggling readers in grades 4–12: 1980–2011*. *Journal of Learning Disabilities*, 48(4), 369–390. <https://doi.org/10.1177/0022219413504995>.
- Schulte, A. C., Stevens, J. J., Elliott, S. N., Tindal, G., & Nese, J. F. T. (2016). *Achievement gaps for students with disabilities: Stable, widening, or narrowing on a state-wide reading comprehension test?*. *Journal of Educational Psychology*, 108(7), 925–942. <https://doi.org/10.1037/edu0000107>.
- Scruggs, T. E., Mastropieri, M. A., & McDuffie, K. A. (2007). *Co-teaching in inclusive classrooms: A meta-synthesis of qualitative research*. *Exceptional Children*, 73(4), 392–416. <https://doi.org/10.1177/001440290707300401>.
- Shaffer, L., & Thomas-Brown, K. (2015). *Enhancing teacher competency through co-teaching and embedded professional development*. *Journal of Education and Training Studies*, 3(3), 117–125. <https://doi.org/10.11114/jets.v3i3.685>.
- * Shaw, F. R. (2002). *Academic achievement of students with disabilities in co-teaching, resource room, and support facilitation models* (Publication No. 3071528) [Doctoral dissertation. Florida Atlantic University]. ProQuest Dissertations & Theses Global. <https://proquest.com/dissertations-theses/academic-achievement-students-with-disabilities/docview/276297958/se-2?accountid=14541>.
- Showalter, D., Klein, R., Johnson, J., & Hartman, S. L. (2017). *Why rural matters*. The Rural School and Community Trust. <https://eric.ed.gov/?id=ED590169>.
- Sindelar, P. T., Pua, D. J., Fisher, T., Peyton, D. J., Brownell, M. T., & Mason-Williams, L. (2018). *The demand for special education teachers in rural school revisited: An update on progress*. *Rural Special Education Quarterly*, 37(1), 12–20. <https://doi.org/10.1177/8756870517749247>.
- Spencer, M., & Wagner, R. K. (2018). *The comprehension problems of children with poor reading comprehension despite adequate decoding: A meta-analysis*. *Review of Educational Research*, 88(3), 366–400. <https://doi.org/10.3102/0034654317749187>.
- Stefanidis, A., King-Sears, M. E., & Brawand, A. (2019). *Benefits for co-teachers of students with disabilities: Do contextual factors matter?. Psychology in the Schools*, 56(4), 539–553. <https://doi.org/10.1002/pits.22207>.
- Stockard, J., Wood, T. W., Coughlin, C., & Khoury, C. R. (2018). *The effectiveness of direct instruction curricula: A meta-analysis of a half century of research*. *Review of Educational Research*, 88(4), 479–507. <https://doi.org/10.3102/0034654317751919>.
- Strogilos, V., Stefanidis, A., & Tragoulia, E. (2016). *Co-teachers' attitudes towards planning and instructional activities for students with disabilities*. *European Journal of Special Needs Education*, 31(3), 344–359. <https://doi.org/10.1080/08856257.2016.1141512>.
- Suurmond, R., van Rhee, H., & Hak, T. (2017). *Introduction, comparison, and validation of meta-essentials: A free and simple tool for meta-analysis*. *Research Synthesis Methods*, 8(4), 537–553. <https://pubmed.ncbi.nlm.nih.gov/28801932/>.
- Sweigart, C. A., Collins, L. W., Evanovich, L. I., & Cook, S. C. (2016). *An evaluation of the evidence base for performance feedback to improve teacher praise using CEC's quality indicators*. *Education & Treatment of Children*, 39(4), 419–444. <https://muse.jhu.edu/article/633657>.
- Szumski, G., Smogorzewska, J., & Karwowski, M. (2017). *Academic achievement of students without special educational needs in inclusive classrooms: A meta-analysis*. *Educational Research Review*, 21(1), 33–54. <https://doi.org/10.1016/j.edurev.2017.02.004>.
- Talbott, E., Maggin, D. M., Van Acker, E. Y., & Kumm, S. (2017). *Quality indicators for reviews of research in special education*. *Exceptionality*, 26(4), 245–265. <https://doi.org/10.1080/09362835.2017.1283625>.
- Therrien, W. J., Taylor, J. C., Hosp, J. L., Kaldenberg, E. R., & Gorsh, J. (2011). *Science instruction for students with learning disabilities: A meta-analysis*. *Learning Disabilities Research & Practice*, 26(4), 188–203. <https://doi.org/10.1111/j.1540-5826.2011.00340.x>.
- * Tremblay, P. (2013). *Comparative outcomes of two instructional models for students with learning disabilities: Inclusion with co-teaching and solo-taught special education*. *Journal of Research in Special Educational Needs*, 13(4), 251–258. <https://doi.org/10.1111/j.1471-3802.2012.01270.x>.
- UNESCO. (2017). *A guide for ensuring inclusion and equity in education*. Paris: UNESCO.
- Unrau, N. J., Rueda, R., Son, E., Polanin, J. R., Lundeen, R. J., & Muraszewski, A. K. (2018). *Can reading self-efficacy be modified? A meta-analysis of the impact of interventions on reading self-efficacy*. *Review of Educational Research*, 88(2), 167–204.
- Valentine, J. C., & Cooper, H. (2008). *A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: The Study Design and Implementation Assessment Device (Study DIAD)*. *Psychological Methods*, 13(2), 130–149. <https://doi.org/10.1037/1082-989X.13.2.130>.
- Van Garderen, D., Stormont, M., & Goel, N. (2012). *Collaboration between general and special educators and student outcomes: A need for more research*. *Psychology in the Schools*, 49(5), 483–497. <https://doi.org/10.1002/pits.21610>.
- * Walker Harris, L. (2009). *Team teaching: The impact on students with disabilities in a middle school setting* (Publication No. 3355861) [Doctoral dissertation. Capella University]. ProQuest Dissertations & Theses Global. <https://proquest.com/docview/305165201?accountid=14541>.
- * Weichel, W. A. (2001). *An analysis of student outcomes in cotaught settings in comparison to other special education service delivery options for students with learning disabilities* (Publication No. 3021407) [Doctoral dissertation. University of California-Riverside]. ProQuest Dissertations & Theses Global. <https://proquest.com/docview/250831425?accountid=14541>.
- Weiss, M. P., & Glaser, H. (2021). *Instruction in co-teaching in the age of Endrew F*. *Behavior Modification*, 45(1), 39–65. <https://doi.org/10.1177/0145445519836071>.
- Wexler, J., Kearns, D. M., Lemons, C. J., Mitchell, M., Clancy, E., Davidson, K. A., ... Wei, Y. (2018). *Reading comprehension and co-teaching practices in middle school English language arts classrooms*. *Exceptional Children*, 84(4), 384–402. <https://doi.org/10.1177/0014402918771543>.

- * Williams, R. L. (2012). *Service delivery models: An effectiveness study in select schools* (Publication No. 3517189) [Doctoral dissertation. Capella University]. ProQuest Dissertations & Theses Global. <https://proquest.com/docview/1030968325?accountid=14541>.
- Woods, A. D. (2020). Explaining longitudinal patterns of special education service receipt. *Exceptional Children*, 87(1), 5–26. <https://doi.org/10.1177/0014402920960655>.
- Zigmond, N. (2015). Where should students with disabilities receive their education?. In B., Bateman, S., Wills-Lloyd, & M., Tankersley (Eds.), *Enduring issues in special education: Personal perspectives* (pp. 198–213). Routledge.
- Zigmond, N., Magiera, K., Simmons, R., & Volonino, V. (2013). Strategies for improving student outcomes in co-taught general education classrooms, In B. G., Cook, & M., Tankersley M. (Eds.), *Research-based strategies for improving outcomes in academics* (pp. 116–124). Pearson.

UNCORRECTED PROOF