Research in Developmental Disabilities 47 (2015) 80–92

Contents lists available at ScienceDirect

Research in Developmental Disabilities

A meta-analysis of the effects of placement on academic and social skill outcome measures of students with disabilities

Conrad Oh-Young *, John Filler

University of Nevada, Las Vegas, Department of Educational & Clinical Studies, 4505 S. Maryland Parkway, Box #453014, Las Vegas, NV 89154-3014, United States

ARTICLE INFO

Article history:
Received 9 February 2015
Received in revised form 24 June 2015
Accepted 13 August 2015
Available online 4 September 2015

Keywords:
Meta-analysis
Inclusion
Integration
Placement
Least restrictive environment
Achievement
Outcomes
Social skills
Social interaction

ABSTRACT

This study involved an investigation of differences between outcome measures of students with disabilities placed in more integrated settings with those of students placed in less integrated settings. A meta-analysis was conducted using the findings from 24 studies published in peer-reviewed journals from 1980 through 2013. Results from the analyses suggest that there were significant differences ($p < 0.0001$) between placement settings with the majority of students with disabilities in more integrated settings outperforming those in less integrated settings on both academic and social outcome measures. Overall these findings, combined with those from two prior meta-analytic studies, provide evidence spanning over 80 years suggesting separate settings are not as beneficial as are more integrated settings. Implications related to practice and policy, as well as avenues for future study, are discussed.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

The Individuals with Disabilities Education Act Regulations (IDEA, 2006) mandate that, "To the maximum extent appropriate, children with disabilities, including children in public or private institutions or other care facilities, are educated with children who are nondisabled" (34 C.F.R. § 300.114 [a] [2] [i]), and that, "Special classes, separate schooling, or other removal of children with disabilities from the regular educational environment occurs only if the nature or severity of the disability is such that education in regular classes with the use of supplementary aids and services cannot be achieved satisfactorily" (34 C.F.R. § 300.114 [a] [2] [ii]). But not all children with disabilities receive their education in the same classrooms or settings as their peers without disabilities. For example, consider 2012 data reported by the U.S. Department of Education (2014) on individuals who received services under Part B of IDEA: (a) 23.6% of the children ages three through five were educated in a separate class; (b) 13.8% of students ages six through twenty-one received special education and related services less than 40% of the time in regular education classrooms; and (c) just 13.1% of individuals identified as having multiple disabilities and 17.1% of individuals identified as having intellectual disabilities received special education and related services in the regular classroom for more than 80% of the

* Corresponding author. Tel.: +1 7028951105.
E-mail addresses: cohyoung@gmail.com (C. Oh-Young), jfiller@unlv.nevada.edu (J. Filler).

http://dx.doi.org/10.1016/j.ridd.2015.08.014
0891-4222/© 2015 Elsevier Ltd. All rights reserved.

day. Clearly there are differences between the amounts of time children with disabilities are educated in the general education setting.

1.1. Differences in placement settings

The prevalent practice of placing some students in general education settings and others in more restrictive separate settings begs the question of efficacy. Are students learning in both settings and are there differences favoring one setting over the other? There have been numerous articles going back many decades arguing for the preference for one type of placement over the other (e.g., Hyatt & Filler, 2011; Sailor et al., 1989; Sindelar & Deno, 1978; Taylor, 1988; Zigmond & Baker, 1996; Zigmond, Kloo, & Volonino, 2009), but far fewer actual data-based comparisons of relative gains made in inclusive versus more restrictive settings have been published in peer-reviewed journals (Wang & Baker, 1985–1986; Zigmond, 2003). Even fewer, Carlberg and Kavale (1980) and Wang and Baker (1985–1986), have attempted a statistical summary analysis of those studies.

1.2. Meta-analysis

A meta-analysis is a statistical procedure that can be used to quantitatively synthesize findings reported across a group of studies (Borenstein, Hedges, Higgins, & Rothstein, 2009; Lipsey & Wilson, 2001). At the very heart of meta-analyses are the calculations of effect sizes from individual studies that are, "... a way of quantifying the size of the difference between two groups" (Coe, 2002, p. 1). The individual effect sizes (Cohen's *d* or Hedges' *g*) are then used to calculate a summary effect, which is essentially an, "... estimate of the mean of these effects" (Borenstein et al., 2009, p. 78).

1.3. Prior analyses

Carlberg and Kavale (1980) performed a meta-analysis to summarize findings from 50 research studies from 1932 into the 1970s. These researchers found that regardless of the types of measures (academic, social, or other) used to assess student performance, placement in the regular classroom led to improved outcomes over placement in the special class environment. This was especially true for students who had IQs that were below average. However, when the researchers separated these results by disability type, they found that the ideal placement setting for students who were diagnosed as having a learning disability, behavioral disorder, or emotional disturbance was the special classroom.

Wang and Baker (1985–1986) performed a meta-analysis of the findings published in journal articles from 1975 through spring of 1984. Once again the finding was that students with disabilities who received instruction in mainstreamed settings outperformed those who received instruction in non-mainstreamed settings. They found no significant differences between overall effect sizes after factoring in such other variables as methodology used, participant grade level, diagnosed disability and level of mainstreaming. They concluded that placement setting seemed to be the variable that most influenced results.

1.4. Purpose

Though it has been argued that determining the differences between placement settings is not an appropriate question to ask (Zigmond, 2003) the fact remains that there are a large number of children with disabilities who continue to receive instruction in settings that limit access to, or simply do not include, their peers without disabilities. The continued prevalent reality of separate settings demands that questions of relative efficacy continue to be asked and variables associated with any differences be examined. Moreover, instructional strategies validated by empirical studies of comparative efficacy have increasingly been applied in both integrated and segregated settings (see Browder, Spooner, Ahlgrim-Delzell, Harris, & Wakeman, 2008; Spooner, Knight, Browder, & Smith, 2012) and so, the differences once noted may no longer be as apparent. Therefore, the purpose of this study is to perform a meta-analysis of findings from peer-reviewed journal articles published between the years of 1980 through 2013 to answer the question: Are there differences between placement of students with disabilities (preschool to high school) in more integrated settings versus placement in less integrated settings using assessments that measure academic and social outcomes reported in the most recent series of studies?

2. Method

2.1. Software used

The software programs EndNote Version 6 for Macintosh OS, Microsoft Excel 2011 for Macintosh OS, and Comprehensive Meta-Analysis Version 2 for Windows 8 were used in this study. EndNote Version 6 was used to create and house the literature database. Excel was used to perform the meta-analytic calculations. Comprehensive Meta-Analysis Version 2 was used to verify calculations and to generate forest plot graphs.

2.2. Literature search

The following search and review procedures were utilized to identify studies for inclusion in the analysis. The search consisted of five different tiers and was adapted using guidelines provided by Moher, Liberati, Tetzlaff, Altman, and The PRISMA Group (2009). The first author worked in conjunction with the second author on each tier. The entire process took over two years to complete.

2.2.1. First and second tiers: Identification and initial screening

A search of the online databases ERIC, Child Development and Adolescent Studies, PsycINFO, and Scopus from 1980 through 2013 was conducted using the terms "inclusion" OR "mainstreaming" OR "LRE" OR "least restrictive environment" OR "integration" AND "disabilities" in abstract, keywords, or title. These words or phrases alone or in any combination returned articles for further review. The start date for the search was set at 1980 for two reasons: (a) because both the Carlberg and Kavale (1980) and Wang and Baker (1985–1986) studies sampled and analyzed the findings from the literature up to the 1980s (in the case of Wang and Baker up to spring of 1984); and (b) the authors of the current study saw little value in re-analyzing data already reported in the literature.

The search for references between the years 1980 through 2013 returned 34,757 entries. This marked the First Tier. Results returned included peer-reviewed journal articles, non-peer-reviewed journal articles, reports, conference submissions, books, book reviews, and dissertations. Initial screening occurred in which the titles and abstracts of the results were reviewed within their respective search engine databases. Entries related to the field of education and individuals with disabilities were then exported from each database into EndNote Version 6 to create a Second Tier set of 10,946 articles. This database is referred to as the EndNote Inclusion Database (EID). Entries that were not related to the field of education or individuals with disabilities were not exported.

2.2.2. Third and fourth tiers: Identification and screening

The parameters that were used to further search the EID for journal articles from the years 1980 through 2013 included: (a) achievement; (b) outcome; (c) social skills; (d) prosocial; and (e) social interaction. For example, a query that was used for the year 2012 and the keyword "achievement" would be: Reference Type IS journal article AND Year IS 2012 AND Any Field CONTAINS achievement. The Reference Type IS "journal article" search setting ensured that the returns included only entries of articles that were published in journals. The Year IS "2012" search setting instructed the EID to return journal entries that were published in 2012. The Any Field CONTAINS "achievement" search setting option returned those 2012 journal entries in which the word "achievement" appeared anywhere within the citation entry such as in the title, abstract, keywords, or notes sections. The Boolean term AND was used to tie together the three search settings. Thus this search example returned only those entries within the database that were listed as journal articles, published in the year 2012, and contained the term "achievement" within the citation entry (e.g., title, abstract, keywords). This procedure was repeated for each year from 1980 through 2013. A total of 1278 references were returned (Third Tier). PDF versions of these were retrieved. In the event that the PDF version was unattainable, a full text version was used in its place. Six were eliminated because they were either duplicate entries or it was not possible to retrieve PDF or full text versions. This brought the total to 1272 (Fourth Tier).

2.2.3. Eligibility and inclusion criteria

At this point both authors reviewed articles with a focus on the abstracts, methods, data analyses, and results sections. The objective was to identify the type of publication (e.g., research, literature review, policy analysis, position statement), research design, description of participants, data reported, and country of origin of its participants. The analysis inclusion and exclusion criteria that were used are summarized in Table 1. A total of 21 studies met the inclusion criteria.

2.2.4. Fifth tier: Additional searches

Two additional searches were conducted. The first search focused on studies that were published in the *Journal of Special Education* and *Exceptional Children* between the years of 1980 through 2013 not already identified in prior searches because of failure to meet search term criteria. Both journals were selected because research studies and reviews related to mainstreaming/inclusion have frequently appeared in these journals. The second additional search involved retrieving studies that were published by any of the researchers of the studies that were already included in the analysis but, again, because of a mismatch in search terms, these may have been overlooked. Searches were conducted for each researcher regardless if they were listed in the first, second, or tertiary positions. Both of these searches yielded additional articles bringing the total to 1522 (Fifth Tier). Studies were reviewed using the criteria in Table 1. Three additional studies met the analysis inclusion criteria and were added to the pool of articles bringing the total to 24. Table 2 contains descriptive information related to these studies.

2.2.5. Quality indicators

Gersten et al. (2005) provided a listing of indicators that researchers, editors, and grant application reviewers could use to review the quality of quantitative experimental and quasi-experimental studies. After presenting the indicators Gersten et al. suggested that, "To be considered high quality, a proposal or study would need to meet all but one [nine out of ten] of

Table 1
Inclusion and exclusion criteria.

Criteria	Description
Inclusion criteria	
Publication criteria	Published between the years 1980 through 2013
	Published in a peer-reviewed journal
	Must not have been included in analyses conducted by either Carlberg and Kavale (1980) or Wang and Baker (1985–1986)
Participant criteria	Participants were identified as individuals with disabilities by the authors of the study
	Participants were between the ages of 3 through 21
	Participants were enrolled in a school in the US
	Participants were students enrolled in a school level (grade) that ranged between preschool through high school
Design criteria	Utilized a quantitative or quasi-experimental group design that compared the performance of one group of participants to another group of participants
Placement criteria	Included at least one group of participants who received instruction in a more integrated setting and at least one group of participants who received instruction in a less integrated setting. A more integrated setting was defined as a setting in which participants with disabilities were provided with increased opportunities to interact with their peers without disabilities. In contrast, a less integrated setting was defined as a setting in which participants with disabilities did not have as many opportunities to interact with their peers without disabilities
Data criteria	Reported the number of research participants in each group, setting in which the experiment took place, and the means, and standard deviations that represented the outcomes related to differences in placement settings. Studies did not have to report an effect size in order to be included in the analysis
	Data reported must have been collected through an assessment (e.g., SSRS, WPPSI-R, WJ-3), through frequency counts, or be the participants' GPA
Exclusion criteria	
Publication criteria	Results were reported in a prior study
Participant criteria	Findings related to participants' lives after leaving high school
Design criteria	SCRD studies
	Within-subject comparisons
Placement criteria	Placement patterns were examined
Data criteria	Unable to separate data reported for participants with disabilities from those without disabilities
	The research participants assessed themselves (e.g., peer sociometric data)
	Unable to determine if social skills data measured involved appropriate or inappropriate interactions
	Did not report sufficient data to calculate an effect size

Note. GPA = Grade point average; SCRD = Single case research design; SSRS = Social Skills Rating System—Teacher Version; WJ-3 = Woodcock-Johnson III Tests of Achievement; WPPSI-R = Wechsler Preschool and Primary Scale of Intelligence-Revised.

the Essential Quality Indicators and demonstrate at least four of the quality indicators listed as Desirable" (Gersten et al., 2005, pp. 152–153). They also suggested that, "To be considered acceptable quality, a research proposal or study would need to meet all but one [nine out of ten] of the Essential Quality Indicators and demonstrate at least one of the quality indicators listed as Desirable" (Gersten et al., 2005, p. 152). Even though the purpose of this study was not to review the quality of the publications, the first and second authors applied the 10 Gersten et al. essential quality indicators to the 24 articles in the analysis with the idea that the summary calculations are directly dependent upon the studies in the analysis. In other words, analysis of data from quality studies may yield more meaningful results while analysis of data from lower quality studies may lead to inaccurate/biased results (Borenstein et al., 2009).

Gersten et al. (2005) separated the essential indicators into four categories: (a) description of participants; (b) description of conditions; (c) description of outcome measures; and (d) description of data analysis techniques and results. Prior to this review the authors of this study decided that each article must meet at least one of the essential quality indicators in each category and at least 5 out of the 10 indicators overall. This was based on the methods, results, and discussion provided in three reviews of the literature that incorporated the quality indicators: Browder et al. (2008), Kalef, Reid, and MacDonald (2013), and Kasner, Reid, and MacDonald (2012), and taking into account Gersten et al.'s caveat that, "These definitions of acceptable and high quality are tentative and should be field-tested by universities, agencies that review grant applications, and research organizations" (Gersten et al., 2005, p. 153).

2.3. Inter-rater reliability

The first author reviewed all of the articles in the Fifth Tier ($N = 1522$). The second author reviewed a randomly selected sample of 305 (20%) in the Fifth Tier. Inter-rater reliability was calculated by dividing the number of studies that the second author excluded from the analysis by the number of studies the first author excluded from the analysis and then multiplying the quotient by 100% (Ayers & Ledford, 2014). A 99% agreement was obtained.

2.4. Data extraction

The first author reviewed the methods, data analyses, and results sections to identify the number of participants, the means, and the standard deviations for all groups. These data were entered into three separate Excel spreadsheets

Table 2
Studies comparing outcomes in more integrated settings against less integrated settings.

Study	Population		Setting type	
	Diagnosis[a]	Age range	More integrated (*N*)	Less integrated (*N*)
Allen and Osborn (1984)	Deafness, EBD/BD, HI, ID, MD, LD, OHI, VI	8-Older than 18	Integrated (418)	Segregated (796)
Cole, Waldron, and Majd (2004)	EBD/BD, ID, LD, OHI	2nd–5th Grade	Integrated (194)	RR (236)
Cole et al. (1991)	DD	3–6	Integrated (42)	SC (58)
Guralnick et al. (1996a)	DD	4.5–5.5	Integrated PG (12)	Segregated PG (18)
Guralnick et al. (1996b)	SLI	4.25–5.5	Integrated PG (12)	Segregated PG (18)
Harris et al. (1990)	Autism	4–5.5	Integrated (5)	Segregated (5)
Holahan and Costenbader (2000)	DD	3–5	Integrated (15)	SCC (15)
Jenkins et al. (1989)	EBD/BD, ID, OHI, PI, SLI	3–6	Integrated (32)	Segregated (21)
Kluwin (1993)	Deafness	Adolescents	Regular MS (35) Some MS (34)	Varied MS (102) SCC (54)
Kluwin and Moores (1985)	Deafness	*M* = 17	Integrated (36)	SCC (44)
Kurth and Mastergeorge (2010)	Autism	12.25–15.75	Integrated (7)	SCC (8)
Lane et al. (2005a)	Deafness, EBD/BD, ID, LD, OHI, SLI	K-8th Grade	SCC (29)	SCS (43)
Lane et al. (2005b)	EBD/BD, ID, LD, OHI, SLI	K-8th Grade	SCC (26)	SCS (34)
Marston (1996)	LD	Elementary	Integrated (33)	RR (36)
Marston and Heistad (1994)	MiD	2nd–6th Grade	Integrated (358)	RR (312)
Mattison (2011)	EBD/BD	*M* Treatment = 12.8 *M* Comparison = 13	SCC (17)	SCS (59)
Meadows, Neel, Scott, and Parker (1994)	BD	*M* = 13.3	Integrated (13)	SCC (6)
Murawski (2006)	LD	*M* = 14	MS (8)	RR (14)
Osborne, Schulte, and McKinney (1991)	LD	*M* = 7.25[b]	Integrated (14)	RR (25)
Rafferty et al. (2003)	NSD, SVD	2.75–4.75	Integrated (68)	SCC (28)
Rea et al. (2002)	LD	*M* Treatment = 14.5 *M* Comparison = 14.7	Integrated (36)	RR (22)
Schulte et al. (1990)	LD	*M* Treatment = 8.3 *M* Comparison = 8.5	RR[d] (19)	RR[e] (15)
Stevens and Slavin (1995)	LD[f]	2nd–6th Grade	Integrated (68) Integrated (72)	RR (53) RR (65)
Waldron and McLeskey (1998)	LD	2nd–6th Grade	Integrated (71)	RR (73)

Note. Only information directly related to this analysis are listed. Information related to other participants (e.g., those without disabilities) were not included. K = Kindergarten. Different placement settings include: MS = Mainstreamed; PG = Play Group; RR = Resource Room/Pull-Out; SCC = Self-Contained Classroom; SCS = Self-Contained School.

[a] Diagnosed types of disabilities include: BD = Behavioral Disorders; DD = Developmentally Delay; EBD/BD = Emotional Behavioral Disorders/Behavioral Disorders; HI = Hearing Impairment; ID = Intellectual Disabilities; LD = Learning Disabilities; MiD = Mild Disabilities; MD = Multiple Disabilities; NSD = Not Severe Disabilities; OHI = Other Health Impairments; PI = Physical Impairments; SVD = Severe Disabilities; SLI = Speech or Language Impairments (listed as Communication Disorders in some studies); VI = Visual Impairment.

[b] Mean reported age is average of all research participants, both those with and without disabilities.

[c] Rea et al. (2002) included only GPA of C or above.

[d] Resource room for one period a day.

[e] Resource room for two periods a day.

[f] Stevens and Slavin (1995) referred to participants with LD but labeled their entire group of participants as "academically handicapped" (p. 249).

along with information related to the authors of the studies, the year each the study was published, and the names of the measures (i.e., assessments) that were used to collect data. Data related to both academic achievement and social skills/social interactions for all research participants with disabilities were included in the first spreadsheet, the Combined Data Set. Examples of the types of measures used to collect academic achievement data included, but were not limited to, the Basic Academic Skills Samples (BASS), WPPSI-R, and the WJ-3. Examples of the types of measures used collect social skills/social interaction data included, but were not limited to, the SSRS and the Preschool Language Scale-3 (PLS-3). Data related to academic achievement for all research participants with disabilities were entered into the second spreadsheet, the Academic Achievement Data Set, and data related to social skills/social interactions for all research participants with disabilities were entered into the third spreadsheet, the Social Data Set.

If the outcomes that were measured fell into both academic and social domains, data were included in both the academic and social data sets. An example of this can be found in Rafferty, Piscitelli, and Boettcher (2003) in the expressive language outcome as measured using the PLS-3. This particular outcome was included in both academic and social domains because that measure was comprised of abilities reflected in both the academic (e.g., "integrative thinking skills") and social domains (e.g., "social communication"; Rafferty et al., 2003, p. 471).

Data entered from participants placed in more integrated settings were labeled and referred to as the treatment group. Data from participants who were placed in less integrated settings were labeled and referred to as the comparison group. For studies that incorporated multiple groups, data from only those groups whose placement settings were most similar to each

other were compared. For example, outcomes from participants in self-contained (SC) and partially integrated settings were compared instead of outcomes from participants in SC and fully integrated settings.

2.5. Data analysis

Effect sizes, the standard errors, variance, and lower and upper confidence interval limits for measures related to all academic and social outcomes were calculated using data entered into the spreadsheets. Due to controversy surrounding the use of Cohen's *d* (Browder et al., 2008), Hedges' *g* effects sizes were also calculated. These results were then used to calculate the summary effect, also known as the standardized mean difference, using a random effects model. The random effects model was selected because it involves the assumption that each study has its own true effect size as a result of the interventions/treatments being conducted by different individuals in different settings at different intensities and because it would be the most appropriate model when attempting to generalize findings to a larger population (Borenstein et al., 2009).

Besides the summary effect the I^2 statistic for the three data sets was calculated. The purpose of the I^2 statistic is to measure the heterogeneity of effects sizes across studies (Borenstein et al., 2009; Higgins, Thompson, Deeks, & Altman, 2003). It is a percentage that ranges from 0 to 100% with a value of 0% representing homogeneity or no variation across studies, 25% representing low heterogeneity across studies, 50% representing moderate heterogeneity across studies, and 75% representing high heterogeneity across studies (Higgins et al., 2003; Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006). The lower the I^2 value, the lower the probability that the differences between effects sizes across studies are the result of errors related to sampling (Huedo-Medina et al., 2006). The higher the I^2 value, the higher the probability that the differences between effects sizes across studies are the result of differences between the studies themselves and not related to sampling error (Huedo-Medina et al., 2006).

All calculations for each data set were first conducted using formulas inputted into Excel then verified using Comprehensive Meta-Analysis Version 2. In a few instances the standard deviation that was reported in the original study was zero. When this occurred, a standard deviation of 0.00000000000000000001 was entered into both programs. This was done to enable the software to properly calculate both *d* and *g* effect sizes (if both standard deviations equaled zero the programs would report that the effect size was undefined).

3. Results

Application of the Gersten et al. (2005) essential quality indicators revealed that all of the studies met at least one indicator in each of the four categories, every study met at least five out of the ten indicators, and two (Lane, Wehby, Little, & Cooley, 2005a; Lane, Wehby, Little, & Cooley, 2005b) met 9 out of the 10. However, none of the studies met all 10 of the essential quality indicators.

3.1. Summary effects

Summary effects for both Cohen's *d* and Hedge's *g* for all three data sets, Combined Data Set, Academic Achievement Data Set, and Social Data Set were first calculated in Excel and then verified in Comprehensive Meta-Analysis Version 2. Forest plots highlighting the results of the analyses are presented in Fig. 1 for Cohen's *d* effect sizes and Fig. 2 for Hedge's *g* effect sizes.

In both figures the value that represents the summary effect (the standardized difference in means) for each data set is indicated by the center of the diamond. For example, if one were to draw a vertical line through the center of the diamond

Data Set	Standardized Difference in Means (*d*)	Standard Error	Variance	Lower Limit	Upper Limit	Z-Value	*p* Value
Combined	0.288	0.033	0.001	0.224	0.352	8.822	< 0.0001
Academic	0.309	0.037	0.001	0.235	0.382	8.244	< 0.0001
Social	0.257	0.059	0.004	0.14	0.373	4.323	< 0.0001

Fig. 1. Summary effects (*d*) all data sets.

Data Set	Standardized Difference in Means (*g*)	Standard Error	Variance	Lower Limit	Upper Limit	Z-Value	*p* Value
Combined	0.284	0.032	0.001	0.221	0.347	8.829	< 0.0001
Academic	0.306	0.037	0.001	0.233	0.378	8.254	< 0.0001
Social	0.251	0.058	0.003	0.137	0.364	4.332	< 0.0001

Fig. 2. Summary effects (*g*) all data sets.

under the Combined Data Set in Fig. 1, one would see that line would be a little over half way between the zero and 0.50 effect size markers (0.288 to be precise). The left most edge of the diamond and the right most edge of the diamond indicate the lower and upper limits (95% CI) that were calculated (0.224 and 0.352).

3.1.1. Combined

Using the results presented across 24 different studies, 141 different effect sizes were calculated for the Combined Data Set for both academic achievement and social skills. The Cohen's d summary effect was $d = 0.29$, $p < 0.0001$, 95% CI [0.22, 0.35] indicating a significant difference between more integrated settings and less integrated settings. The Hedges' g summary effect was $g = 0.28$, $p < 0.0001$, 95% CI [0.22, 0.35] once again indicating a significant difference between more integrated settings and less integrated settings. Utilizing the effect size interpretation scale presented in Coe (2002), these results suggest that if the data were normally distributed, that approximately 62% of the individuals in less integrated settings performed below the average performance of individuals in more integrated settings on measures that assessed academic and/or social outcomes.

Out of the 141 effect sizes in this analysis, 81 were solely related to academic outcomes, 50 solely related to social outcomes, and 10 were related to both academic and social outcomes (and thus reflected across all analyses). Effect sizes for individual measures within studies ranged from $d = -1.07$ to $d = 4.40$ ($M = 0.35$, $SD = 0.56$, $Mdn = 0.27$) and $g = -1.04$ to $g = 4.14$ ($M = 0.34$, $SD = 0.54$, $Mdn = 0.25$).

A total of 29 effect sizes were less than zero suggesting that on these measures participants in less integrated settings out performed those individuals who were in more integrated settings. These results represented 20.57% of the overall sample included in this analysis and were almost evenly divided between outcome types with 15 solely related to academic outcomes, 13 solely related to social outcomes, and 1 related to both academic and social outcomes.

There were four effect sizes that were equal to zero suggesting that the difference in setting did not have a significant effect on these outcomes. One outcome was academic related and three were social related. This accounted for 2.84% of the sample included in this analysis.

Finally, a total of 108 effect sizes, 76.6% of the overall sample included in this analysis, were above zero suggesting that on these measures, participants in more integrated settings out performed those individuals in less integrated settings. Majority of these outcomes, 65 in total, were solely related to academic outcomes. Thirty-four of the remaining outcomes were related to social outcomes and nine fell into both academic and social domains.

3.1.2. Academic achievement

Summary effects related to the Academic Achievement Data Set were calculated using data presented in 24 studies and led to the calculation of 91 effect sizes. The Cohen's d summary effect was $d = 0.31$, $p < 0.0001$, 95% CI [0.24, 0.38] indicating a significant difference between more integrated settings and less integrated settings. The Hedges' g summary effect was $g = 0.31$, $p < 0.0001$, 95% CI [0.23, 0.38] once again indicating a significant difference between the two settings. If normally distributed, according to Coe (2002), approximately 62% of those individuals in less integrated settings performed below the average performance of those individuals in more integrated settings on these measures.

Of the 91 effect sizes that comprised the Academic Achievement Data Set, a total of 16 (17.58%) were less than zero, 1 (1.1%) was equal to zero, and 74 (81.32%) were above zero. Individual effect sizes ranged from $d = -0.67$ to $d = 4.40$ ($M = 0.39$, $SD = 0.58$, $Mdn = 0.28$) and $g = -0.65$ to $g = 4.14$ ($M = 0.38$, $SD = 0.56$, $Mdn = 0.28$). There were no academic outcomes with an effect size less than negative one. There were six outcomes with an effect size greater than one.

3.1.3. Social interaction

The Social Data Set summary effects were calculated using data across 8 studies resulting in the calculation of 60 effects sizes. The Cohen's d summary effect was $d = 0.26$, $p < 0.0001$, 95% CI [0.14, 0.37] indicating a significant difference between more integrated settings and less integrated settings when comparing social outcomes. Individual effect sizes ranged from $d = -1.07$ to $d = 1.79$ ($M = 0.28$, $SD = 0.48$, $Mdn = 0.22$). The Hedges' g summary effect was $g = 0.25$, $p < 0.0001$, 95% CI [0.14, 0.36] with individual effect sizes ranging from $g = -1.04$ to $g = 1.75$ ($M = 0.27$, $SD = 0.46$, $Mdn = 0.21$). This indicated that once again there was a significant difference between settings. If normally distributed, these data, according to Coe (2002), would suggest that majority of the individuals in less integrated settings, between 58% and 62%, performed below the mean performance of individuals in more integrated settings on measures related to social outcomes.

Of the 60 effect sizes in this analysis, a total of 14 effect sizes (23.33%) were less than zero, 3 (5%) were equal to zero, and 43 (71.67%) were above zero. Majority of the effect sizes, 69.64%, were between 0 and 1. One effect size was less than negative one and three were greater than one.

3.2. Heterogeneity

Heterogeneity for the Combined Data Set was $I^2 = 59\%$ suggesting that there exists substantial heterogeneity across treatment effects across all studies in the analysis. Heterogeneity for the Academic Achievement Data Set was found to be $I^2 = 65\%$ once again suggesting substantial heterogeneity. Finally heterogeneity for the Social Data Set was calculated to be $I^2 = 39\%$. This suggests little to moderate heterogeneity.

3.3. Study comparisons

A post-hoc analysis was conducted to determine if the findings from this study were significantly different from those reported in the earlier Carlberg and Kavale (1980) study. In order to perform the analysis mean effect sizes using Glass's Δ were calculated using the data in this study. Three independent *t*-tests were then performed. Results suggest that there were significant differences between the combined outcomes $t(461) = 2.93$, $p < 0.005$ and academic outcomes $t(216) = 2.21$, $p < 0.05$ between the two studies. There was no significant difference between the social outcomes data $t(219) = 1.26$, $p = 0.21$. Regardless of the significant differences between the combined and academic outcomes, all six mean effect sizes favored placement in more integrated settings.

4. Discussion

The issue of placement setting for students with disabilities has been frequently debated in the literature (Hyatt & Filler, 2011; Sailor et al., 1989; Sindelar & Deno, 1978; Taylor, 1988; Zigmond & Baker, 1996; Zigmond et al., 2009). Studies providing quantitative evidence favoring one placement setting over another have been published less frequently (Wang & Baker, 1985–1986; Zigmond, 2003). Findings published over 20 years ago have suggested that the optimal educational placement setting for students with disabilities is the more integrated setting (Carlberg & Kavale, 1980; Wang & Baker, 1985–1986). However, researchers since then have conducted studies using instructional strategies that have been shown to be effective in both settings (e.g., Browder et al., 2008; Spooner et al., 2012).

The purpose of this study was to investigate if there were differences between placement of students with disabilities in more integrated settings versus placement in less integrated settings on academic and social outcomes given recent advances in instructional strategy. A meta-analysis was performed to synthesize findings across 24 studies from peer-reviewed journal articles published between the years of 1980 through 2013. Results suggest that there exist significant differences between placement settings with participants in more integrated settings outperforming those in less integrated settings.

4.1. Individualizing results

The authors sought to address this study's research question from a broad perspective hence the analysis of data from participants who were diagnosed with different types of disabilities from different age groups who were placed in different settings with the idea that this would enhance the generalizability of findings. This involved utilizing meta-analytic procedures to synthesize findings across a collection of studies (Borenstein et al., 2009). However, results from a meta-analysis may also yield findings beyond the summary effect (Borenstein et al., 2009).

4.1.1. Exclusion

There is a notion that students with more severe disabilities are placed in less integrated, more restrictive settings based on the severity of their disabilities (Lane et al., 2005a, 2005b; Rafferty et al., 2003). While these findings do not directly address that frame of thought, they do provide evidence that relates to that notion on both a macro level and a micro level. At the macro level, this study's findings clearly suggest that the majority of participants placed in more integrated settings outperformed their peers in less integrated settings. This dispersion was continuous across all three data sets. On a micro level, consider the findings from Kurth and Mastergeorge (2010) where participant WJ-3 scores ranged from 52 to 85 for those participants who received instruction in an inclusive setting and 1–47 for those who received instruction in a SC (self-contained) setting. The lowest score in the inclusive setting was greater than the highest score in the SC setting with only one individual in the SC setting achieving a WJ-3 score greater than 23. These results suggest one of two possibilities: (1) there existed significant differences between groups; or (2) the impact of the differences in settings was very large. An inspection of the IQ scores between groups revealed no significant differences between participants in both settings leading Kurth and Mastergeorge to conclude that while all of the participants in their study had the ability to perform academic tasks, those who received instruction in the SC setting lacked the skills necessary to perform those tasks.

Overall, these findings suggest that if educators are segregating students based solely on severity of disability then we, at the very least, may be providing a disservice to a population that needs our support. We may be denying them access to settings in which, with the appropriate services and supports, they may be able to achieve more than they could otherwise.

4.1.2. Severity of disability

The fact that individuals with severe disabilities may be segregated based on the level of severity raises the question: Do these individuals benefit from being placed in more integrated settings? Prior literature reviewed by Zigmond et al. (2009) suggested that the more severe the disability, the more likely that these students would experience a curriculum that is less academically focused. For example, the findings in Kluwin (1993) suggested that students with hearing impairments who received instruction in more integrated settings were more likely to be exposed to a program that was academic focused while those in more segregated settings were likely to be exposed to a program that was vocationally focused. Unfortunately the majority of studies in this analysis did not separate results by severity of disability making it difficult to delineate differences within the population. Two studies that did, Rafferty et al. (2003) and Waldron and McLeskey (1998), did so to varying degrees.

In Rafferty et al. (2003), the researchers grouped participants based on WPPSI-R performance and then reported their scores separately (severe versus not severe). An inspection of the academic and social outcomes results suggested that overall, participants with severe disabilities benefited from placement in inclusive settings more so than did their peers not classified as having severe disabilities. Placement had little to no effect for participants who did not have severe disabilities as evaluated by the SSRS ($d = 0.10$; $g = 0.10$). In contrast, differences in placement for participants with severe disabilities as measured by the SSRS were more apparent ($d = 0.94$; $g = 0.92$). Rafferty et al. found that these differences were further represented in expressive language ($d = 0.84$; $g = 0.83$) and auditory comprehension ($d = 0.80$; $g = 0.79$) outcomes measured by the PLS-3 for participants with severe disabilities versus expressive language ($d = -0.05$; $g = -0.05$) and auditory comprehension ($d = 0.28$; $g = 0.28$) outcomes for those not with severe disabilities.

In Waldron and McLeskey (1998), the researchers separated participants into severe and mild LD categories based on performance on the Kaufman Test of Educational Achievement Brief Form. Unfortunately their data were reported in aggregate (both mild and severe combined), therefore the results of the analysis were from the combined group. However, and of interest to this discussion, Waldron and McLeskey compared progress made by participants in both LD categories (severe and mild) placed in both settings (inclusive and segregated) to progress made by students without disabilities on reading and math outcomes measured by the BASS. The researchers reported in narrative that they found a significant difference between progress made by participants with mild LD who were educated in inclusive settings when compared to progress made by students with mild LD in segregated settings on reading outcomes. No significant differences were found on the other measures.

4.1.3. Instructional support

Since the two prior meta-analytic studies were published, researchers have identified indicators of quality research conducted in the field (e.g., Gersten et al., 2005; Horner et al., 2005). Using these indicators, evidence has been collected regarding the effective use of instructional strategies across a variety of settings (see Browder et al., 2008; Spooner et al., 2012). Wang and Baker (1985–1986) may have hinted at the effects of these supports when referring to an upward trend that favored placement in more integrated settings in articles published after 1970. Indeed, results from the post-hoc analysis conducted in this study support this notion with significant differences between the combined and academic outcomes calculated in Carlberg and Kavale (1980) and the combined and academic outcomes calculated in this study.

An inspection of the procedures used within studies in this analysis suggested that majority identified participants in separate settings and then measured if there were differences between participant outcomes after a given amount of time had passed. For example, in Guralnick, Connor, Hammond, Gottman, and Kinnish (1996b) the researchers separated participants into MS (mainstream) playgroups and segregated playgroups, exposed them to typical preschool classroom conditions without any extra support, and found that participants in MS playgroups engaged in solitary play more frequently than those in segregated playgroups.

However, a handful of studies involved participants who received instruction in different placement settings and were exposed to instructional strategies that varied depending on the setting. Two such studies in which this occurred were Schulte, Osborne, and McKinney (1990) and Stevens and Slavin (1995). Four effect sizes were calculated from Schulte et al. (1990) which ranged from $d = 0.48$ to $d = 1.02$ and $g = 0.09$ to $g = 1.00$ and eight effect sizes were calculated from Stevens and Slavin (1995) which ranged from $d = 0.09$ to $d = 0.40$ and $g = 0.09$ to $g = 0.40$. All 12 suggested differences between placement settings with participants in more integrated settings outperforming their peers.

Overall, these findings further support the commonly held belief that placement in more integrated settings, with exposure to the appropriate instructional strategies and supports, such as those highlighted in Browder et al. (2008) and Spooner et al. (2012), can only lead to improved student outcomes. It is, after all, what makes special education special (Cook & Schirmer, 2003).

4.1.4. Quality indicators

Application of the Gersten et al. (2005) quality indicators to past studies revealed mixed results and suggested the need for caution when drawing conclusions. The suggestion put forth that studies should meet all but one of their essential indicators in order to be considered "acceptable" and of "high quality" (Gersten et al., 2005, p. 152) sets a laudable but extremely high standard. In the 14 studies identified in Browder et al. (2008), none could be considered to be high quality by Gersten et al. standards. Findings from Kalef et al. (2013) indicated that one out of five studies reviewed could be considered high quality. Finally, results from Kasner et al. (2012) suggested that out of the two studies reviewed, both were not considered high quality. In this study 2 out of the 24 studies in the analysis met the high quality essential quality indicator requirement. While we must continue to improve the quality of research by whatever standards we may chose we are still left with less than perfection and must draw what conclusions we can with full knowledge of the need for caution.

One major area of scrutiny seems to be centered on the quality indicator that suggests that researchers report effect sizes. In all three-research reviews (Browder et al., 2008; Kalef et al., 2013; Kasner et al., 2012) majority of the studies, 20 out of 22 did not include effect size calculations. In this study 10 out of the 24 studies reported an effect size of some type (e.g., Cohen's d, Hedges' g, Glass's Δ, correlation coefficient [r], eta-squared [η^2]). Note that we are not suggesting that effect sizes should not be reported. However, it should be noted that whether or not researchers calculate effect sizes is immaterial when performing a meta-analysis because the process of performing the analysis itself results in the calculation of individual effect sizes for each study. Moreover, what matters is that each of the individual studies included in the analysis provide enough

data and description to allow for missing information to be derived after the fact. Thus, it is of little surprise that individuals in the field have suggested that the quality indicators should be revisited and/or revised (Kalef et al., 2013; Kasner et al., 2012) and that further discussion should occur especially when involving participants with low incidence disabilities (Browder et al., 2008).

4.2. Limitations

4.2.1. Publication bias

Publication bias, sometimes referred to as the "file drawer problem", refers to the notion that studies that are included within a meta-analysis tend to be those that have been published (Borenstein et al., 2009; Lipsey & Wilson, 2001). The findings related to the analysis, therefore, will be biased toward the findings within those studies (Borenstein et al., 2009), hence the upward trend favoring more integrated settings (Wang & Baker, 1985–1986). With this in mind it is important to remember that publication bias is not limited only to meta-analytic studies (Borenstein et al., 2009). Publication bias affects all reviews of the literature (Borenstein et al., 2009) because many of the findings that are analyzed are from peer-reviewed journals. In addition, the purpose of this analysis was to analyze those data published in peer-reviewed journals between 1980 through 2013. Therefore, our population of studies was all peer-reviewed journal articles published from 1980 through 2013. The sample represents a subset of that population obtained by using four different search engine databases, keyword searches, and analysis inclusion and exclusion criteria.

4.2.2. Heterogeneity

Heterogeneity results for this study ranged from little or none to moderate (Social Data Set) to substantial heterogeneity (Combined and Academic Achievement Data Sets). This may suggest that there were inherent differences between the studies that were not related to sampling error. One possible cause for this difference could be because of the constructs that were measured. The majority of the outcomes in the current analysis were related to academic achievement (81 out of 141). Effect sizes varied widely across different studies, for example reading performance in Guralnick, Connor, Hammond, Gottman, and Kinnish (1996a) where $d = -0.53$ and $g = -0.52$ and in Schulte et al. (1990) where $d = 1.02$ and $g = 1.00$, across different content areas, such as reading performance in Rea, McLaughlin, and Walther-Thomas (2002) where $d = 0.24$ and $g = 0.23$ and math performance in Lane et al. (2005b) where $d = 0.86$ and $g = 0.85$, and different age groups such as the differences in reading performance across grade levels in Marston and Heistad (1994) where $d = -0.26$ and $g = -0.26$ for second grade students and $d = 0.40$ and $g = 0.40$ for sixth grade students.

Furthermore, this does not take into consideration the performance outcomes of individuals diagnosed with different types of disabilities. For example, consider that there were studies in this analysis that only included participants with low incidence disabilities (e.g., Kluwin & Moores, 1985; Kurth & Mastergeorge, 2010), there were studies that only included participants with high incidence disabilities (e.g., Murawski, 2006; Rea et al., 2002), and there were studies that included individuals from both groups (e.g., Allen & Osborn, 1984; Rafferty et al., 2003). In addition, the types of disabilities across studies ranged from health related impairments to behavioral disorders. These individuals represented different age groups, were placed across different settings at different levels of integration (SC to full integration), received instruction from different individuals at different intensities, and received different services and supports. All of these represent variables that may have exerted their influences on the results.

4.2.3. Practical issues

Practical issues related to sampling may have had an impact on the results. For example, random assignment of participants was suggested as an essential quality indicator by Gersten et al. (2005) and was implemented in a select number of studies (see Cole, Mills, Dale, & Jenkins, 1991; Jenkins, Odom, & Speltz, 1989; Schulte et al., 1990). However, the educational setting that a given study takes place in may make it difficult for researchers to utilize random assignment (Gersten et al., 2005; Harris, Handleman, Kristoff, Bass, & Gordon, 1990), much less obtain access to sufficient numbers of participants. Therefore, inability to randomly assign across all studies, not to mention randomly select participants, may have impacted these findings.

4.3. Implications

4.3.1. Implications for practice

To parents attending an Individualized Education Program (IEP) meeting a Cohen's d effect size of 0.29 may not mean much. However, what that value means is that data collected through research suggest that: (a) their child's academic and social outcomes could be related to placement setting (more integrated setting versus less integrated setting); and that (b) placement in a more integrated setting may lead to improved outcomes over a less integrated setting.

It is also important to consider that researchers in a handful of studies did more than separate participants into integrated and segregated groups and then measure the differences between (e.g., Schulte et al., 1990; Stevens & Slavin, 1995). These studies involved different instructional strategies. It is the quality of these instructional strategies, along with evidence-based practices such as those described in Browder et al. (2008) and Spooner et al. (2012), that teachers, teacher educators,

and administrators need to be cognizant of. As summarized by Rafferty et al. (2003), "Simply stated, for inclusion to be successful, program quality must be high and appropriate support services must be provided" (p. 478).

4.3.2. Implications for policy

To place the overall findings into perspective one must consider the argument provided by Hedges and Hedberg (2007) that even though Cohen (1988) suggests that an effect size of 0.2 demonstrates a small effect, this result still holds meaning in the field of education. The reason being that a summary effect of zero would have indicated that there was no difference between the different groups suggesting there was no benefit from being exposed to the treatment. However, it is important to remember that, "Some tiny effects may save lives or have theoretical importance; some large effects are of little interest" (Cumming, Fidler, Kalinowski, & Lai, 2012, p. 140). Small changes, such as an intervention implemented correctly that results in a small effect might translate to substantial benefits over time (Coe, 2002).

4.4. Future research

Although the purpose of this study was to examine the differences placement may have on students with disabilities using data published between the years of 1980 through 2013, further study is needed in at least three areas. The first is related to how different settings impact students with different types and degrees of disabilities. For example, previous findings (Carlberg & Kavale, 1980) suggested differences between more and less integrated placement settings for students with high-incidence disabilities. Is there any suggestion that placements in more integrated settings for students with low-incidence disabilities differ from placements in less integrated settings? Furthermore, are there differences between placement settings when comparing students with high-incidence disabilities and students with low-incidence disabilities? Might the differential effect of setting vary across age groups (elementary, middle, and high school)? Finally, might the level of integration: (a) full integration in the general education setting; (b) 50% of instructional time in the general education setting; (c) SCC in a regular school; or (d) SCS, have an effect on the results?

A second area for future attention involves expansion of setting comparisons to include studies that were conducted outside of the U.S. The U.S. is clearly not the only country where students with disabilities are educated in integrated and segregated settings. Including these studies will serve to improve both the validity and the generalizability of findings.

Yet a third consideration involves expanding the search for literature. Publication bias notwithstanding, the search for studies could be expanded through the use of additional keywords (e.g., impairment and developmental delay) and the use of wildcards (e.g., impair*). Taking publication bias into account, future reviews could include findings that were not published in peer-reviewed journals (e.g., dissertations, conference presentations, and reports), something not done in this study. However, expansion along any of these lines will take an incredible amount of time and resources. The pool of references returned using the search terms in this study was 34,757. It took over two years of regular work to synthesize and analyze this large data set.

5. Conclusion

More than 60 years have passed since the U.S. Supreme Court rendered the decision in Brown v. Board of Education (1954). Since then many individuals have played a role in shaping policy and practice such that children with disabilities are now able to gain access to and receive a public education that is both appropriate and delivered in the LRE. But as long as there are children who receive their education in environments that do not include their peers without disabilities we must continue to ask questions regarding the efficacy of separate placement. Placement is, or should be, determined by the needs of each child. If those needs are not being met, as determined by measures of outcome, then the appropriateness of the placement should be questioned.

Results from this study suggest that there were significant differences between more integrated settings and less integrated settings. Furthermore, results suggest that overall, students who received instruction in more integrated settings outperformed students who received instruction in less integrated settings on assessments that measured academic and social outcomes. In combination with the results presented by Carlberg and Kavale (1980) and Wang and Baker (1985–1986), these findings provide evidence spanning over 80 years that reiterate and reinforce the notion that separate is not always equal.

References

Allen, T. E., & Osborn, T. I. (1984). Academic integration of hearing-impaired students: Demographic, handicapping, and achievement factors. *American Annals of the Deaf, 129*(2), 100–113. http://dx.doi.org/10.1353/aad.2012.1529. The study is included in the meta-analysis

Ayers, K., & Ledford, J. R. (2014). Dependent measures and measurement systems. In D. L. Gast & J. R. Ledford (Eds.), *Single case research methodology: Applications in special education and behavioral sciences* (2nd ed., pp. 124–153). New York, NY: Routledge.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. New York, NY: Wiley http://dx.doi.org/10.1002/9780470743386

Browder, D. M., Spooner, F., Ahlgrim-Delzell, L., Harris, A. A., & Wakeman, S. (2008). A meta-analysis on teaching mathematics to students with significant cognitive disabilities. *Exceptional Children, 74*(4), 407–432. http://dx.doi.org/10.1177/001440290807400401

Brown v. Board of Education (1954). 347 U.S. 483; 74 S. Ct. 686; 98 L. Ed. 873. In *U.S. LEXIS 2094; 53 Ohio Op. 326; 38 A.L.R.2d 1180*. Brown v. Board of Education Retrieved from ⟨www.lexisnexis.com/hottopics/lnacademic⟩.

Carlberg, C., & Kavale, K. (1980). The efficacy of special versus regular class placement for exceptional children: A meta-analysis. *Journal of Special Education, 14*(3), 295–309. http://dx.doi.org/10.1177/002246698001400304
Coe, R. (2002 September). It's the effect size, stupid: What effect size is and why it is important. In *Paper presented at the annual conference of the British Educational Research Association*. England: University of Exeter.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
Cole, C. M., Waldron, N., & Majd, M. (2004). Academic progress of students across inclusive and traditional settings. *Mental Retardation, 42*(2), 136–144. http://dx.doi.org/10.1352/0047-6765(2004)42<136:APOSAI>2.0.CO;2. The study is included in the meta-analysis
Cole, K. N., Mills, P. E., Dale, P. S., & Jenkins, J. R. (1991). Effects of preschool integration for children with disabilities. *Exceptional Children, 58*(1), 36–45. The study is included in the meta-analysis
Comprehensive Meta-Analysis (Version 2.2.064) [Computer software]. Englewood, NJ: Biostat.
Cook, B. G., & Schirmer, B. R. (2003). What is special about special education?. Overview and analysis. *The Journal of Special Education, 37*(3), 200–205. http://dx.doi.org/10.1177/00224669030370031001
Cumming, G., Fidler, F., Kalinowski, P., & Lai, J. (2012). The statistical recommendations of the American psychological association publication manual: Effect sizes, confidence intervals, and meta-analysis. *Australian Journal of Psychology, 64*(3), 138–146. http://dx.doi.org/10.1111/j.1742-9536.2011.00037.x
EndNote (Version 6) [Computer software]. Philadelphia, PA: Thomson Reuters.
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. http://dx.doi.org/10.1177/001440290507100202
Guralnick, M. J., Connor, R. T., Hammond, M., Gottman, J. M., & Kinnish, K. (1996). Immediate effects of mainstreamed settings on the social interactions and social integration of preschool children. *American Journal on Mental Retardation, 100*(4), 359–377. The study is included in the meta-analysis
Guralnick, M. J., Connor, R. T., Hammond, M. A., Gottman, J. M., & Kinnish, K. (1996). The peer relations of preschool children with communication disorders. *Child Development, 67*(2), 471–489. http://dx.doi.org/10.2307/1131827. The study is included in the meta-analysis
Harris, S. L., Handleman, J. S., Kristoff, B., Bass, L., & Gordon, R. (1990). Changes in language development among autistic and peer children in segregated and integrated preschool settings. *Journal of Autism and Developmental Disorders, 20*(1), 23–31. http://dx.doi.org/10.1007/BF02206854. The study is included in the meta-analysis
Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis, 29*(1), 60–87. http://dx.doi.org/10.3102/0162373707299706
Higgins, J. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal, 327*(7414), 557–560. http://dx.doi.org/10.1136/bmj.327.7414.557
Holahan, A., & Costenbader, V. (2000). A comparison of developmental gains for preschool children with disabilities in inclusive and self-contained classrooms. *Topics in Early Childhood Special Education, 20*(4), 224–235. http://dx.doi.org/10.1177/027112140002000403. The study is included in the meta-analysis
Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165–179. http://dx.doi.org/10.1177/001440290507100203
Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods, 11*(2), 193–206. http://dx.doi.org/10.1037/1082-989X.11.2.193
Hyatt, K. J., & Filler, J. (2011). LRE re-examined: Misinterpretations and unintended consequences. *International Journal of Inclusive Education, 15*(9), 1031–1045.
Individuals with Disabilities Education Act Regulations (2006). Retrieved from http://idea.ed.gov/download/finalregulations.pdf.
Jenkins, J. R., Odom, S. L., & Speltz, M. L. (1989). Effects of social integration on preschool children with handicaps. *Exceptional Children, 55*(5), 420–428. The study is included in the meta-analysis
Kalef, L., Reid, G., & MacDonald, C. (2013). Evidence-based practice: A quality indicator analysis of peer-tutoring in adapted physical education. *Research in Developmental Disabilities, 34*(9), 2514–2522. http://dx.doi.org/10.1016/j.ridd.2013.05.004
Kasner, M., Reid, G., & MacDonald, C. (2012). Evidence-based practice: Quality indicator analysis of antecedent exercise in autism spectrum disorders. *Research in Autism Spectrum Disorders, 6*(4), 1418–1425. http://dx.doi.org/10.1016/j.rasd.2012.02.001
Kluwin, T. N. (1993). Cumulative effects of mainstreaming on the achievement of deaf adolescents. *Exceptional Children, 60*(1), 73–81. The study is included in the meta-analysis
Kluwin, T. N., & Moores, D. F. (1985). The effects of integration on the mathematics achievement of hearing impaired adolescents. *Exceptional Children, 52*(2), 153–160. The study is included in the meta-analysis
Kurth, J. A., & Mastergeorge, A. M. (2010). Academic and cognitive profiles of students with autism: Implications for classroom practice and placement. *International Journal of Special Education, 25*(2), 8–14. The study is included in the meta-analysis
Lane, K. L., Wehby, J. H., Little, M., & Cooley, C. (2005a). Academic, social, and behavioral profiles of students with emotional and behavioral disorders educated in self-contained classrooms and self-contained schools: Part I—Are they more alike than different? *Behavioral Disorders, 30*(4), 349–361. The study is included in the meta-analysis
Lane, K. L., Wehby, J. H., Little, M., & Cooley, C. (2005b). Students educated in self-contained classrooms and self-contained schools: Part II—How do they progress over time? *Behavioral Disorders, 30*(4), 363–374. The study is included in the meta-analysis
Lipsey, M. W., & Wilson, D. V. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.
Marston, D. (1996). A comparison of inclusion only, pull-out only, and combined service models for students with mild disabilities. *Journal of Special Education, 30*(2), 121–132. http://dx.doi.org/10.1177/002246699603000201. The study is included in the meta-analysis
Marston, D., & Heistad, D. (1994). Assessing collaborative inclusion as an effective model for the delivery of special education services. *Diagnostique, 19*(4), 51–67. http://dx.doi.org/10.1177/073724779401900404. The study is included in the meta-analysis
Mattison, R. E. (2011). Comparison of students classified ED in self-contained classrooms and a self-contained school. *Education & Treatment of Children, 34*(1), 15–33. http://dx.doi.org/10.1353/etc.2011.0003. The study is included in the meta-analysis
Meadows, N. B., Neel, R. S., Scott, C. M., & Parker, G. (1994). Academic performance, social competence, and mainstream accommodations: A look at mainstreamed and nonmainstreamed students with serious behavioral disorders. *Behavioral Disorders, 19*(3), 170–180. The study is included in the meta-analysis
Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(6), e1000097. http://dx.doi.org/10.1371/journal.pmed1000097
Murawski, W. W. (2006). Student outcomes in co-taught secondary English classes: How can we improve? *Reading & Writing Quarterly: Overcoming Learning Difficulties, 22*(3), 227–247. http://dx.doi.org/10.1080/10573560500455703. The study is included in the meta-analysis
Osborne, S. S., Schulte, A. C., & McKinney, J. D. (1991). A longitudinal study of students with learning disabilities in mainstream and resource programs. *Exceptionality, 2*(2), 81–95. http://dx.doi.org/10.1080/09362839109524769. The study is included in the meta-analysis
Rafferty, Y., Piscitelli, V., & Boettcher, C. (2003). The impact of inclusion on language development and social competence among preschoolers with disabilities. *Exceptional Children, 69*(4), 467–479. http://dx.doi.org/10.1177/001440290306900405. The study is included in the meta-analysis
Rea, P. J., McLaughlin, V. L., & Walther-Thomas, C. (2002). Outcomes for students with learning disabilities in inclusive and pullout programs. *Exceptional Children, 68*(2), 203–222. http://dx.doi.org/10.1177/001440290206800204. The study is included in the meta-analysis
Sailor, W. et al. (1989). *The comprehensive local school: Regular education for all students with disabilities*. Baltimore, MD: Paul H Brookes Publishing Co.
Schulte, A. C., Osborne, S. S., & McKinney, J. D. (1990). Academic outcomes for students with learning disabilities in consultation and resource programs. *Exceptional Children, 57*(2), 162–172. The study is included in the meta-analysis

Sindelar, P. T., & Deno, S. L. (1978). The effectiveness of resource programming. *The Journal of Special Education, 12*(1), 17–28. http://dx.doi.org/10.1177/002246697801200105

Spooner, F., Knight, V. F., Browder, D. M., & Smith, B. R. (2012). Evidence-based practice for teaching academics to students with severe developmental disabilities. *Remedial and Special Education, 33*(6), 374–387. http://dx.doi.org/10.1177/0741932511421634

Stevens, R. J., & Slavin, R. E. (1995). Effects of a cooperative learning approach in reading and writing on academically handicapped and nonhandicapped students. *The Elementary School Journal, 95*(3), 241–262. http://dx.doi.org/10.1086/461801. The study is included in the meta-analysis

Taylor, S. J. (1988). Caught in the continuum: A critical analysis of the principle of the least restrictive environment. *Journal of the Association for Persons with Severe Handicaps, 13*(1), 41–53.

U.S. Department of Education, Office of Special Education and Rehabilitative Services, Office of Special Education Programs (2014). *36th Annual Report to congress on the implementation of the Individuals with Disabilities Education Act, 2014*. Washington, DC.

Waldron, N. L., & McLeskey, J. (1998). The effects of an inclusive school program on students with mild and severe learning disabilities. *Exceptional Children, 64*(3), 395–405. http://dx.doi.org/10.1177/001440299806400308. The study is included in the meta-analysis

Wang, M. C., & Baker, E. T. (1985–1986). Mainstreaming programs: Design features and effects. *The Journal of Special Education, 19*(4), 503–521. http://dx.doi.org/10.1177/002246698501900412

Zigmond, N. (2003). Where should students with disabilities receive special education services? Is one place better than another?. *The Journal of Special Education, 37*(3), 193–199. http://dx.doi.org/10.1177/00224669030370030901

Zigmond, N., & Baker, J. M. (1996). Full inclusion for students with learning disabilities; too much of a good thing? *Theory into Practice, 35*(1), 26–34. http://dx.doi.org/10.1080/00405849609543698

Zigmond, N., Kloo, A., & Volonino, V. (2009). What, where, and how? Special education in the climate of full inclusion. *Exceptionality, 17*(4), 189–204. http://dx.doi.org/10.1080/09362830903231986