Race and ethnicity in education intervention research: A systematic review and recommendations for sampling, reporting, and analysis

Larissa M. Gaias[1*]

Mylien T. Duong[2]

Michael D. Pullmann[1]

Stephanie K. Brewer[1]

Michelle Smilansky[1]

Mallory Halbert[1]

Cathea M. Carey[1]

Janine Jones[3]

[1]School Mental Health Assessment, Research, and Training Center
University of Washington School of Medicine

[2]Committee for Children

[3] University of Washington College of Education

*Corresponding author: Please direct any inquiries or concerns regarding this manuscript to Larissa Gaias, University of Washington SMART Center, 6200 NE 47th Street , Seattle, WA 98115; lgaias@uw.edu, (631) 375-1849.

Declarations of interest: None

Funding: This work was supported by the Institute of Education Sciences [grant number: R305B170021]

Publication Date: November 2020, Educational Research Review

Abstract

Racial/ethnic educational disparities remain a pervasive and intractable issue facing the U.S. education system. To eliminate these disparities, educational research must consistently attend to race/ethnicity, particularly when examining the effects of educational practices, programs, and policies. The goal of the current review was to examine the prevalence with which educational intervention research 1) reports on race/ethnicity, 2) includes samples representative of the racial/ethnic composition of U.S. public schools, and 3) examines the potential for reducing racial/ethnic disparities. We coded a randomly selected 96 (13%) of all educational intervention studies that met What Works Clearinghouse (WWC) standards for rigorous research in the areas of students with disabilities, literacy, early childhood education, English Language Learners, mathematics, character education, and dropout prevention. We also identified and coded 210 meta-analyses in these topic areas. Because our aim was to describe the extent to ``reporting, sampling, and analysis, we included studies even when reducing disparities was not an explicit focus. Results indicated that 27% of empirical studies and 94% of meta-analyses did not report race/ethnicity. Hispanic/Latinx, Asian/Pacific Islander, and Native American youth were underrepresented in research samples. Only 19% of empirical studies and 6% of meta-analyses conducted analyses to examine the potential impact for reducing racial/ethnic disparities. This review demonstrates a clear need for researchers to be more intentional when designing and evaluating educational interventions to reduce pervasive racial/ethnic inequities in educational outcomes.

Race and ethnicity in educational intervention research: A systematic review and recommendations for sampling, reporting, and analysis

Educational inequity is a serious global concern, with both within- and between-country disparities impacting who has access to high-quality education that is responsive to students' academic, social-emotional, and behavioral needs, and facilitates long-term health and well-being throughout the lifespan (Hadjar & Gross, 2016). Although educational inequity occurs on multiple levels in many, if not all, international settings, the nature of inequity within any given context is often shaped by factors specific to that context, such as laws and historical events that affect race relations (Hadjar & Gross, 2016). For example, educational inequities within the United States (US) have been impacted by longstanding legacies of colonialism and colonization, slavery, immigration, segregation & desegregation, and gentrification (Bertocchi & Dimico, 2012; DeSena & Ansalone, 2009; Gonzales, Heredia, & Negrón-Gonzales, 2015; Orfield & Eaton, 1996; Pewewardy, 2000; Seawright, 2014; Stillman, 2012). Considering the historical and current manifestations of systemic bias and racism within US education systems, many scholars, educational leaders, and advocates have emphasized the need to identify strategies that reduce pervasive and persistent racial/ethnic inequities within the country; thus, the US is a particularly relevant context within which to examine the potential for educational interventions to reduce disparities.

Within the US, Black/African American, Hispanic/Latinx, and Native American students have substantially lower test scores, grades, attendance, school engagement, graduation rates, and college enrollment. For example, by the end of the 8th grade, Black/African American, Hispanic/Latinx, and Native American students' reading and mathematics skills are three years behind those of White students (National Center for Education Statistics, 2014). Disparities are

not only present with regard to academic outcomes; marginalized students, particularly Black/African American students, are significantly more likely to be overrepresented in exclusionary disciplinary actions such as office referrals, suspensions, and expulsions, contributing to lost instructional time, lower levels of school belonging, school dropout (Skiba et al., 2002). Students of color also tend to receive fewer and lower quality social-emotional, mental health, and behavioral supports, that are less responsive to their needs and backgrounds (Castro-Olivo, 2017). Addressing the problem of racial/ethnic inequality in education is especially urgent considering that the majority of students enrolled in the US are non-White (National Center for Education Statistics, 2017).

Public schooling has potential to be the "great equalizer" by providing opportunities for all students to develop positive self-concept, self-improvement, and self-empowerment. In an ideal world, educational programs and practices would correct for other systemic inequities. In practice, however, individual and institutional biases within educational systems also perpetuate racial divides in other domains of life. School dropout, for example, predicts later unemployment, lower lifetime earnings and wealth, worse physical and mental health, and greater receipt of social services (Caspi et al., 1998; Reingold & Liu, 2009). Punitive disciplinary practices have often served to funnel marginalized youth out of educational systems into the juvenile and criminal justice systems, through a process commonly referred to as the school-to-prison pipeline (Skiba et al., 2014). Economically, the cost of racial/ethnic disparities in education on the U.S. Gross Domestic Product is estimated to be equivalent to the impact of an economic recession (Auguste et al., 2009).

To address these longstanding inequities, educational researchers need to consistently and systematically attend to race and ethnicity in educational research. This is particularly true for

research that aims to develop programs, practices, and policies that presumably aim to improve student outcomes. These innovations provide a key opportunity to reduce disparities across a range of educational outcomes and domains. Without consistent and systematic attention to issues of equity, educational researchers risk developing and disseminating "evidence-based" programs that could in fact worsen existing disparities. Indeed, educational practitioners often find that recommended curricula are not relevant or responsive to their student populations, thus requiring adaptations to better meet student needs (Toppel, 2013). Superintendents of large urban school districts rate the racial/ethnic achievement gap as one of their major concerns (Huang et al., 2003). These educators have asked for practical, evidence-based advice for what to implement in their schools to address this challenge.

It is unclear, however, the extent to which current educational research is poised to address this identified need of educators and educational decision makers. In order to better inform research and practice regarding the effectiveness of educational interventions for students from marginalized racial identities, researchers need to address pervasive concerns regarding how to (1) consistently and thoroughly report the race/ethnicity of their sample in research publications, (2) include racially and ethnically representative participant samples, and (3) explicitly examine the potential for educational interventions to reduce, maintain, or exacerbate educational inequities. The aim of the current study is to examine the extent to which current research on educational interventions addresses these three domains. Toward this aim, we review educational intervention research, regardless of whether the intent of the original research was explicitly focused on reducing racial/ethnic inequity. Our goal in this study is to provide insight into the state of educational research regarding the prevalence of reporting on race/ethnicity, the inclusion of representative samples, and the examination of disparity-reducing effects.

1.1 Consistent Reporting of Race/Ethnicity

Adequately collecting and reporting race and ethnicity data is essential for communicating information regarding the generalizability of findings and the implications of an intervention for particular racial/ethnic groups. However, there have been longstanding criticisms of the quality with which race and ethnicity data have been reported in the educational literature. For instance, the Council for Learning Disabilities (CLD) research committee raised concern about inconsistently reported demographics in 1984 (Smith et al., 1984) and the National Joint Committee on Learning Disabilities (NJCLD; 1991) called for investigators to report more specific demographic data in 1991. However, a CLD investigation from 1992 found that little progress had been made in the preceding 10 years (Rosenberg et al., 1992). A meta-analysis on 26 reading interventions, conducted in 2012, found that reporting had improved somewhat after the CLD and NJCLD calls for greater reporting quality, but 27% of the studies in this analysis still did not provide a breakdown of racial/ethnic characteristics of their sample (Reed et al., 2013). A recent review of special education intervention studies published between 2010 and 2016 found that only 54.7% of studies reported race/ethnicity (Sinclair et al., 2018).

Improving the collection and reporting of race and ethnicity data has also been encouraged by US federal policies and recommendations. In 1997, the Office of Management and Budget (OMB) issued a revised set of recommendations for the collection and use of race and ethnicity data by federal agencies (Office of Management and Budget, 1997). These recommendations were adopted by the US Department of Education (DOE) in 2007, to be enacted within schools by the 2011-2011 school year (US Department of Education, 2008). This directive recommended that race and ethnicity information be collected by self-report via two questions. The first question asks about ethnicity (i.e., whether a participant is Hispanic or

Latinx) and the second question asks the participant to identify his/her race or races, according to give categories (i.e., American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, Black, White). This directive rejected the use of a "multiracial" category, but recommended that the census allow respondents to select more than one category. In this directive, the DOE also required elementary, secondary, and postsecondary institutions to report student data according to the above racial and ethnic categories.

Some agencies have expanded these recommendations to research projects. For example, the FDA recommended the use of the standardized OMB race and ethnicity categories for data collection in clinical trials (FDA Office of Minority Health, 2016) to facilitate scientific progress. This is currently not a requirement for research projects funded through the DOE, however. In addition, it is unclear whether these reporting practices are applied consistently in research not funded by federal agencies. Some federal organizations have recognized limitations of the OMB directive and have encouraged data collection and reporting standards that better elucidate heterogeneity within racial and ethnic groups and a growing mixed-raced or mixed-ethnicity population. For example, the Office of Minority Health within the US Department of Health and Human Services (2018) recommends asking about more specific identities (e.g., Cuban; Vietnamese) within each of the broad racial/ethnic categories (e.g., Hispanic/Latinx, Asian). Although these standards provide more specific and informative information regarding a participant's race/ethnicity, they may not be applied consistently across research studies.

1.2 Representative Participant Samples

Unfortunately, research as an enterprise has often been plagued by under-representation of non-White racial and ethnic groups. Indeed, some studies have found that research participants represent only a very narrow subset of human populations (Arnett, 2008; Henrich et

al., 2010). Collecting data from non-representative samples seriously limits generalizability and researchers' ability to apply knowledge and findings across a diverse range of ethnic/racial groups. This is especially important in intervention research, where programs and practices are likely to be used in real-world settings with a wide spectrum of demographic groups.

Research conducted in educational settings may be particularly well-suited to collect data from more representative samples. Collecting data in schools allows researchers to engage participants in a setting where they already are, reducing the burden on students or families to transport their children to university laboratories or other research settings (Alibali & Nathan, 2010). Conducting research in schools, however, does not guarantee a diverse and representative sample. Schools are largely segregated along racial lines (Wagner, 2017); in fact, recent research has demonstrated that school segregation is higher than residential segregation and has been increasing over the past two decades, as federal desegregation mandates have been lifted (Orfield, 2013; Orfield & Frankenberg, 2014; Reardon et al., 2012; Sohoni & Saporito, 2009). Thus, researchers need to be intentional in recruiting schools whose demographic characteristics align with the intended sample. In addition, consent rates often differ significantly by race. In one study that was conducted in five diverse regions across the US, 64% of White, but only 52% of Black/African American and 50% of Hispanic/Latinx families, provided active consent for their children to participate in a survey-based research study in their school (Esbensen et al., 1999). Thus, researchers must anticipate the barriers that may prevent students from marginalized groups from providing parental consent or assent to participate in studies, including language differences, homelessness, or distrust of educational institutions and research (Blom-Hoffman et al., 2009; Quinn et al., 2012). In sum, although conducting educational research may facilitate the participation of more racially diverse samples than other social science fields,

researchers must be intentional to achieve adequate inclusion of under-represented participants (e.g., Roosa et al., 2008).

1.3 Effects of Interventions on Disparities

Recruiting representative samples and reporting on race/ethnicity alone does not provide information regarding the effectiveness of an educational intervention for reducing racial and ethnic disparities. In addition to recruiting a diverse sample and reporting sample demographics, researchers need to conduct analyses that specifically investigate the impact of interventions on disparities. Because of the longstanding and pervasive nature of racial/ethnic inequities in education, we argue that issues of equity are always relevant in education research, regardless of topic area (e.g., literacy, mathematics, behavior). Examining the effectiveness of interventions in general, without investigating the effectiveness of the intervention for mitigating racial/ethnic disparities, can disguise inequities in these interventions. Researchers may inadvertently perpetuate the assumption that research findings apply to all students when they do not explicitly test this assumption. In the medical field, only 7% of randomized control trials conducted analyses to test treatment effectiveness according to race/ethnicity, indicating a significant missed opportunity to examine the extent to which underrepresented groups respond to common therapies that are considered to be effective (Gabler et al., 2009).

A primary method for testing an intervention's potential for reducing disparities involves conducting subgroup analyses. This method involves disaggregating intervention results by race/ethnicity to understand whether the program has differential impact for any one racial or ethnic group relative to another. Disparities can only be reduced when interventions have stronger effects for marginalized students (e.g., Bradshaw et al., 2018). In contrast, although interventions that demonstrate equivalent effects across groups have the potential to improve

learning and other outcomes for all students regardless of racial/ethnic background, they may serve to maintain disparities (e.g., Gregory et al., 2018). Finally, interventions that have stronger impacts for non-marginalized students can widen achievement gaps and other educational disparities (e.g., Vincent & Tobin, 2011), even if these interventions are effective overall. A similar approach to stratifying analyses is conducting interaction or moderator analyses, which folds the subgroup analyses into an omnibus testing procedure by including interaction terms (e.g., race x treatment condition).

An additional method for examining an intervention's impact on inequality is to use pre-established statistical norms, benchmarks, or other studies' results to interpret findings from studies that primarily sampled marginalized racial/ethnic groups. Effect sizes in such studies can be compared to a known metric, such as the magnitude of a particular achievement gap (e.g., Green, Blasik, Hartshorn, & Shatten-Jones, 2000). If the intervention effect size is equivalent to or exceeds the established reference metric, the intervention may have disparity-reducing effects. Alternatively, studies can compare the change of growth in an intervention group that primarily or entirely contains individuals from marginalized racial/ethnic identities to a control condition, a group of students from non-marginalized racial/ethnic backgrounds who are not receiving the intervention (e.g., Dobbie & Fryer, 2009). In this case, although the intervention may theoretically be equally effective for marginalized and non-marginalized students, it in fact may have a gap-reducing impact if marginalized students are overrepresented in the intervention condition. Thus, these interventions may reduce disparities by increasing access to resources to a level nearing that of non-marginalized students. Testing for single group effects is important because these studies often assess interventions designed to respond to the particular needs and backgrounds of a marginalized racial group, which may be most effective for improving

outcomes (Williams & Deutsch, 2016).

1.4 The Current Study

The aim of this study was to describe the current treatment of race and ethnicity in educational intervention research. Specifically, we address several research questions:

1) To what extent does educational intervention research adequately report on race/ethnicity? What study characteristics are associated with adequate reporting of the sample race/ethnicity in educational intervention research?
2) To what extent are racial/ethnic groups included and represented in educational intervention research samples?
3) To what extent does educational intervention research examine the potential for reducing racial/ethnic disparities? What study characteristics are associated with conducting analyses to examine disparity-reducing effects?

To answer these research questions, we conducted two systematic reviews. The first was a review of empirical studies reporting results of educational interventions across a wide range of educational domains (e.g., literacy, mathematics, dropout prevention). The second was a review of meta-analyses aligned with those domains. Because our focus was on educational interventions as a whole, we include studies and meta-analyses in our review even if the study's aim was not explicitly focused on reducing inequities. This allows us to draw conclusions regarding the extent to which current educational intervention research can speak to (1) for which populations an intervention may be appropriate or effective, (2) the generalizability of intervention results to the US population, and (3) the possibility for educational interventions to reduce disparities and enhance equity.

2. Methods

2.1 Inclusion Criteria and Selection Procedures

2.1.1 Empirical studies. To be included in the review of empirical studies, studies had to meet the following criteria: (a) were included in the US DOE's What Works Clearinghouse (WWC) database, (b) met WWC standards of rigor with or without reservation (described below) at the time of the review, and (c) were retrievable for coding. The WWC reviews research on different programs, products, practices, and policies in education, with the goal of providing educators with information necessary to make evidence-based decisions. Details on this process can be found in the What Works Clearinghouse Procedures and Standards Handbook, available at http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=19. The WWC conducts reviews of interventions in areas determined by the Institute of Education Sciences (IES) to be of highest priority for informing the US national education policy agenda. Thus, although the WWC is compiled and organized in a manner to best serve the needs of educators within the US, studies that are conducted outside the US are included if they are aligned with the review priorities and criteria. In addition, the findings of the WWC can certainly hold relevance outside the US context. At the time of the current review, the WWC included reviews on seven domains or topic areas: students with disabilities, literacy, early childhood education, English Language Learners, mathematics, character education, and dropout prevention. The search criteria used by WWC can be found in Appendix 1. Databases of both published and unpublished studies ("grey literature") were searched. Although the review process continues to be updated, 10,611 studies had been reviewed by the start of the current study in October 2017.

For the current review, we only included studies that met WWC design standards with or without reservations at the time of review. WWC codes study rigor using a number of criteria, including design, attrition rates, and reliability and validity of outcome measurements. Studies

that meet WWC design standards without reservations include well-implemented randomized controlled trials (RCTs) with no attrition problems. Studies that meet WWC standards with reservations are either (a) RCTs that have some problems with attrition or other design characteristics, or (b) strong quasi-experimental designs. Importantly, design standards are determined independent of study results. That is, a study can meet design standards without having statistically significant intervention outcomes, and vice versa. Of the 10,611 studies reviewed, the WWC deemed that 785 met standards either with or without reservations. Of the 785 studies meeting standards, we randomly selected 100 studies for coding. The full text was available for 96 (12.2%) of these studies, which were included in the current analysis.

2.1.2 Meta-analyses. A bibliographic search was conducted in the following databases: Educational Resource Information Center, PsycArticles, PsycINFO, Social Services Abstracts, Sociological Abstracts, and ProQuest Dissertation and Theses. Consistent with WWC protocol for empirical studies, meta-analyses had to be published within the last 20 years and be written in English. Narrative reviews and empirical reports of original research were not included. Peer-reviewed journal articles, dissertations/theses, books or book chapters, government reports, and conference proceedings were all eligible. We modeled our search procedures after the WWC's published review protocols. For each topic (students with disabilities, literacy, early childhood education, English Language Learners, mathematics, character education, and dropout prevention), we used the WWC published search terms related to the outcome, combined with search terms related to "meta-analysis" (see Appendix 1). Using the WWC protocol, our search yielded meta-analyses that examined the effects of interventions and identified mechanisms for change that can be leveraged via interventions. For instance, a meta-analysis on the impact of parental involvement in schools and student academic functioning included studies of

experiments to increase parental involvement, as well as nonexperimental studies examining the association between parental involvement and student academics. Search results were combined, and duplicates were automatically removed. This yielded 413 records. After screening the titles and abstracts, 175 were deemed ineligible because they were not meta-analyses, another 22 were determined to be duplicates, and we could not obtain the full documentation for an additional 6. This process yielded a final sample of 210 meta-analyses.

2.2 Data extraction and coding

Coding had already been completed by the WWC for some study characteristics, including sample size, participant grade levels, sample gender composition, sample racial/ethnic composition, topic area, percent eligible for free/reduced lunch (FRL), percent eligible for Individual Education Plans (IEP), and percent English Language Learners (ELL). However, some modifications or additional codes were necessary to address our research questions.

2.2.1 Reporting. Studies provided data regarding race/ethnicity with a varying degree of quality. We coded studies as "0" if they reported no race/ethnicity data, "1" for poor quality reporting such as reporting only one race, combining WWC racial categories, or if the total percent of all racial/ethnic groups added to less than 100% (e.g. "the study population was 61% African American and 20% Asian" with no further description of any other race), and "2" if it reported multiple racial categories that added to close to 100% (to allow for rounding errors). For inferential analyses, if a study had any partial or complete information (e.g. a "1" or "2" above), we coded this study as having reported sample race/ethnicity.

2.2.2 Representation. Sample race/ethnicity was coded by the WWC using two categories: ethnicity (percent Hispanic/Latinx[1] and percent not Hispanic/Latinx), and race

[1]Although the original WWC only used the terms "Hispanic" and "Black," we refer to these groups using "Hispanic/Latinx" and "Black/African American," due to the common use of these phrases in academic literature.

(Asian, Black/African American[1], Native American, Pacific Islander, White, Other). Ethnicity and race categories generally combined to 100%, meaning that no categories existed for "Hispanic/Latinx-White" and "Hispanic/Latinx-Black/African American," for example. Our investigation of primary studies found little to no documentation about how multiracial/multiethnic participants were coded (for instance, was a participant who was Hispanic/Latinx and Black/African American coded as Black/African American, Hispanic/Latinx, multiracial, or other?). Therefore, primary studies were coded according to percent Asian, Black/African American, Native American, Pacific Islander, White, Hispanic/Latinx, and Other. Because studies were not consistent with their use of multiracial/multiethnic categorization (and because WWC did not use a multiracial/multiethnic code), the "Other" category includes multiracial students, as well as those who did not identify as any of the categories provided (e.g., Middle Eastern). Only two empirical studies specifically mentioned the inclusion of Pacific Islander students, and many studies collapsed this category with Asian. No meta-analyses reported Pacific Islander separately. Therefore, for these two empirical studies, we combined Asian and Pacific Islander into one category for further analysis.

2.2.3 Testing effects on disparities. The WWC database did not provide coded information about whether a study conducted subgroup analyses by race/ethnicity. A code of "1" was given if a study examined differential outcome effects (i.e., provided disaggregated outcomes by race/ethnicity *or* tested race/ethnicity as a moderator of effects), regardless of the significance of the finding. We also applied a code of "1" if authors discussed the outcomes of such analyses, even if no additional methodological details were provided. We did not include studies that only included race/ethnicity as a covariate or examined mean level differences at baseline, because these do not examine a moderating impact of race on outcomes. We also coded

a study as "1" if it compared effect sizes to external benchmarks to examine the potential impact for reducing disparities. As a point of comparison, we also coded an additional variable that indicated whether studies disaggregated intervention outcomes by socioeconomic status (SES).

2.2.4 Coding reliability. A consensus process was used to develop the coding manual, which began with all coders (two faculty members, who served as PIs for the project, and two undergraduate research assistants) independently coding a random subset of five manuscripts. Coders then met to identify and discuss discrepancies and clarify coding manual instructions. This process was completed three times prior to reaching a consensus on the manual. Raters then proceeded to code the remaining manuscripts independently. A subset of 20% of samples was randomly selected to calculate inter-rater reliability. Inter-rater reliability was calculated with intraclass correlation coefficients (ICCs) for continuous variables and with Fleiss's Kappa for categorical variables. Coders achieved perfect reliability, with 100% agreement.

2.3 Data Analysis Plan

We conducted initial data screening and cleaning of the two datasets (empirical studies and meta-analyses), including calculating descriptive statistics (frequencies/percentages or means/standard deviations) on all study variables.

2.3.1 Reporting. Our first research question examined the extent to which empirical studies and meta-analyses adequately report participant race/ethnicity and which study characteristics are associated with this reporting. We first identified the number of studies with no, partial, or full reporting on race/ethnicity. Then, for empirical studies, we conducted cross-tabulations with χ^2 tests examining the association between our categorical outcome, whether the manuscript had complete, partial, or full reporting on race/ethnicity, and several categorical study characteristics: sample size (divided into quartiles due to skewed distribution and to

improve interpretability); grade level (dummy coded as Pre-K, Elementary [K-5th], Middle [6th – 8th], High [9th-12th], and Post-Secondary); publication type (journal, dissertation/thesis, government report, organizational report); and whether the manuscript reported sample rate of FRL eligibility, ELLs, or students with IEPs. For categorical variables with exclusive response options (i.e., sample size, publication type), we ran an omnibus χ^2 test and, if significant, followed this with individual χ^2 tests for each response option, to identify which response option deviated from the expected distribution. We then computed point-biserial correlations between the reporting variable and our two continuous predictors: publication year and sample rate of FRL eligibility. Similarly, for meta-analyses, we computed cross-tabulations with χ^2 tests between the reporting variable and the number of studies included in the meta-analysis (divided into quartiles due to skewed distribution and to improve interpretability), grade level, publication type, and whether the manuscript included tests of race/ethnicity and/or SES as a moderator. We computed point-biserial correlations between the reporting variable and publication year.

Due to the high number of tests with individual predictors, we conducted Benjamini-Hochberg false discovery rate (FDR) corrections, with the FDR set at 25%, across omnibus tests to mitigate concerns regarding the possibility of Type I error (Benjamini & Hochberg, 1995). The FDR correction is a less conservative alternative to other familywise corrections such as Bonferroni's, as it emphasizes avoiding Type II errors (false negatives) in addition to avoiding Type I errors (false positives). FDR was deemed appropriate for this study, as the impact of a Type I error was of less concern than the impact of a Type II error, considering the low rates of true positives. In the examination of important study characteristics related to race reporting and analysis, false negatives might result in missed opportunities to make recommendations to improve reporting practices.

2.3.2 Representation. Our second research question examined the extent to which each racial/ethnic group that is included in the WWC standard reporting (i.e., Asian/Pacific Islander, Black/African American, Native American, Hispanic/Latinx, White, Other) was represented across empirical studies and meta-analyses of educational interventions. For each race/ethnicity category, we computed the group mean percentage (e.g. mean of the study means) to obtain a "study average" estimate of the representation of racial groups in research in any particular research study. We compared these statistics to the overall percentage of public school-attending youth in 2015 (US Department of Education, 2017), by calculating an odds ratio that reflects the odds that a youth of a particular racial/ethnic group will be included in an empirical study or meta-analysis relative to the prevalence of that group in the population. We computed odds ratios using the formula $\text{OR} = \frac{p2 \,/\, (1 - p2)}{p1 \,/\, (1 - p1)}$, where p_1 is the proportion the racial/ethnic category in the population of youth attending US public schools and p_2 is the grand mean proportion of the racial/ethnic category in the reviewed empirical studies or meta-analyses.

2.3.3 Testing effects on disparities. Our third research question focused on the extent to which empirical studies and meta-analyses conducted analyses to examine the potential for reducing disparities and which study characteristics are associated with these analyses. We first identified the number of empirical studies and meta-analyses that presented such analyses (e.g., disaggregated results by racial groups, included race as a moderator, compared effects to a known benchmark). Then, identical to our analyses for RQ1, we examined the association of study characteristics with whether the study reported these analyses. For empirical studies, we conducted cross-tabulations with χ^2 tests to examine the association between our categorical predictors (i.e., sample size, grade level, publication type, whether the manuscript reported the sample rate of Free or Reduced Lunch, English Language Learners, and students with

Individualized Education Plans) and whether analyses were conducted to examine effects on disparities. Additionally, we computed point-biserial correlations to examine the associations between our two continuous predictors (i.e., publication year, sample rate of Free or Reduced Lunch) and our outcome. Similarly, for meta-analyses, we computed χ^2 tests for categorical predictors (i.e., number of studies included in the meta-analyses, grade level, publication type, whether the manuscript reported race/ethnicity, whether SES was included as a moderator), and point-biserial correlations for our continuous predictor (i.e., publication year). We also conducted a Benjamini-Hochberg FDR correction for family-wise (Type 1) error rate. Finally, for those empirical studies and meta-analyses that did conduct such analyses, we describe these results, indicating whether the intervention demonstrated evidence of reducing, exacerbating, or maintaining disparities.

3. Results

3.1 Descriptive Statistics of Reviewed Studies

3.1.1 Empirical studies. Table 1 depicts the overall descriptive statistics of reviewed empirical studies. The studies took place most commonly in elementary (47.9%) and middle school (32.3%). Study sample size ranged widely, from 10 to 15,661 students, with an average of 1,259, though this was positively skewed by a few large studies (Skew = 3.8, Kurtosis = 16.5). Nineteen studies (19.8%) reported the percent of English Language Learners, and 32 (33.3%) reported the percent of students eligible for free and reduced lunch. Twelve papers (12.5%) analyzed SES as a moderator or stratified analysis by SES. Studies were most commonly published in academic journals (47.9%) or as organizational reports (32.3%). A list of all empirical studies, their study characteristics and their scoring on each construct of interest can be found in Appendix 2.

3.1.2 Meta-analyses. Table 1 also depicts the overall descriptive statistics of reviewed meta-analyses. The 210 meta-analyses included 9,069 individual studies (an unknown number of these studies were duplicates, included in more than one meta-analysis). Most meta-analyses were published in academic journals (n = 174, 82.9%) and addressed character education (n = 67, 31.9%), followed by literacy (n = 56, 26.7%), and math (n = 50, 23.8%). Original research completed in elementary schools were most often included in the 210 meta-analyses (n = 155, 31.3%), followed by middle (n = 138, 27.9%), and high schools (n = 114, 23.0%). A list of all meta-analyses, their study characteristics and their scoring on each construct of interest can be found in Appendix 3.

3.2 Research Question 1: To what extent does educational intervention research adequately report on race/ethnicity? What study characteristics are associated with adequate reporting of sample race/ethnicity in educational intervention research?

3.2.1 Empirical studies. Of the 96 education intervention studies randomly selected for review, 26 (27.1%) provided no detail on race/ethnicity of the sample, 10 (10.4%) reported partial detail (e.g., “the sample was 60% nonwhite”), and 60 (62.5%) reported complete detail on race and ethnicity (see Table 1). Table 2 depicts the associations between study characteristics and whether the study reported on race/ethnicity. Dissertations and studies with fewer than 65 cases, and studies that did not report sample sizes were less likely to report race/ethnicity. Studies with more than 1,273 cases, governmental reports studies that reported FRL, and studies that examined the effectiveness of a study for reducing disparities were significantly were more likely to report race/ethnicity. Publication date was also a significant predictor of reporting quality, with newer studies more likely to report race/ethnicity.

3.2.2 Meta-analyses. Of 210 meta-analyses, 197 (93.8%) provided no detail on the racial

and ethnic characteristics of the sample, 5 (2.4%) provided partial detail, and 8 (3.8%) provided full detail (see Table 1). Table 3 depicts the associations between characteristics of the meta-analysi=s and whether the meta-analysis reported race/ethnicity. Meta-analyses with middle and high school students were significantly less likely to report race/ethnicity. Omnibus tests examining differences according to number of included studies and publication type revealed no significant associations between these variables and reporting practices. Year of publication was also not a significant predictor of reporting practices.

3.3. Research Question 2: To what extent are racial/ethnic groups included and represented in educational intervention research samples?

3.3.1 Empirical studies. Table 4 depicts study racial and ethnic demographics of the reviewed empirical studies and meta-analyses, as well as the composition of public-school attending youth in the US. White youth were most frequently involved in research. They were included in 82.9% of studies that reported at least partial race/ethnicity, constituted 39.9% of the average study sample, and were underrepresented when compared to U.S. public school-attending population estimates (49%, group mean percentage odds ratio [OR] = .69). This OR means that the odds that a White youth will be included in a study is .69 (or 31%) lower than their prevalence in the population. African American students were the next most frequently involved racial/ethnic group. They were included in 80.0% of studies that reported race/ethnicity, constituted 34.9% of the average study sample, and were overrepresented as compared to school attending youth in the US (15%, OR = 3.04). Hispanic/Latinx youth were included in 69.5% of studies that reported race/ethnicity, constituted 13.8% of the average study sample, and were underrepresented as compared to school attending youth (26%, OR = 0.46). Asian/Pacific Islander youth were included in 32.9% of studies that reported race/ethnicity, constituted 2.9% of

the average study sample, and were underrepresented as compared to school attending youth (5%; OR = .57). Native American youth were included in 15.7% of studies that reported race/ethnicity, constituted 0.6% of the average study sample, and were underrepresented in research as compared to their population as school attending youth (1%, OR = .60). Youth classified as "other" race were included in 44.3% of studies that reported race/ethnicity, constituted 6.4% of the average study sample, and were overrepresented in research as compared to the school age population estimates (3%, OR = 2.21).

3.3.2 Meta-analyses. Of the 13 meta-analyses reviewed that reported any race/ethnicity (Table 4), 11 (84.6%) reported including White participants, representing 41.9% of the sample on average, and were underrepresented as compared to school attending youth in the US (OR = 0.75). Black/African American participants were also included in 11 (84.6%) of studies, representing 34.5% of the sample on average, and were overrepresented (OR = 2.98). Hispanic/Latinx participants were reported in 7 studies (53.8%), representing 7.7% of the sample on average, and were underrepresented (OR = 0.23). Asian/Pacific Islander participants were reported in 2 studies (15.4%) representing .5% of the sample on average, and were underrepresented (OR = .10). Native American participants were not reported in any studies. There were 3 (23.1%) studies that reported "other" race, representing 2.1% of the sample on average, and were underrepresented (OR = 0.69). There were 8 studies (61.5%) reporting that at least some of the participants in the population included in the meta-analysis were missing data on race/ethnicity (M = 13.9%, SD = 15.0%). These group mean percentages and their representativeness were similar to the results described in the empirical studies.

3.4 Research Question 3: To what extent does educational intervention research examine the potential for reducing racial/ethnic disparities? What study characteristics are

associated with conducting analyses to examine disparity-reducing effects?

3.4.1 Empirical studies. Of the 96 education intervention studies, 18 (18.8%) conducted analyses that provide information regarding potential effects for reducing educational disparities (see Table 2). Studies that reported race/ethnicity and reported percentage of ELL were significantly more likely to examine effects for reducing disparities.

Of the 18 studies that examined disparity-reducing effects (Table 5), 15 conducted subgroup analyses or disaggregated findings by race; the other three compared single group effects to a benchmark. Among the 15 studies that conducted subgroup analyses, six found that results were equivalent across racial groups (Cave et al., 1993; Gleason et al., 2010; B. W. Hall & Bacon, 2006; Pyke et al., 2004; Smith et al., 1984; Stevens & Durkin, 1992). Four studies found their interventions were most effective for non-White students; these included a social-belongingness intervention for college freshman (Walton & Cohen, 2011), an early childhood (preschool – 1st grade) mathematics intervention (Sarama et al., 2012), a middle school math program (Resendez et al., 2005), and an adolescent civic engagement intervention (Jastrzab, 1997). One study, which examined a values affirmation intervention in undergraduate biology classes, was more effective for White students than non-White students (Harackiewicz et al., 2014). Another study, which evaluated an elementary school literacy intervention, contributed to larger gains in program outcomes for non-Hispanic, but not Hispanic, students (Gunn et al., 2000). Finally, three studies found mixed results. One study evaluating the Milwaukee parental choice program found that intervention effects for Asian and Native American students did not significantly differ from those of White students, but intervention effects were significantly weaker for Hispanic/Latinx and Black/African American students than White students (Witte et al., 2010). A study focused on enhancing preadolescents' coping skills found equivalent

intervention effects for White and non-White students on delinquency, but the program was only effective for White students in relation to substance use (Lochman & Wells, 2004). Finally, an evaluation of an elementary and middle school reading intervention found that race was not a significant moderator of program effects in Grades 3-5, 7, and 8, but the intervention was strongest for White students in sixth grade (Overbay & Baenen, 2003).

Of the three studies that compared single group effects to a benchmark, two found evidence of their intervention for reducing disparities. In one study evaluating the impact of a universal behavior management program for families and parents of pre-school children, intervention effect sizes approached or exceeded the magnitude of known effect sizes (Brotman et al., 2013). In a second study, disparities in standardized math and ELA scores for students in a comprehensive elementary and middle school programs, who primarily identified as Black/African American or Hispanic/Latinx, as compared to a matched sample of White students were closed or cut in half by the end of the program (Dobbie & Fryer, 2009). The third study, however, found mixed results. In this evaluation of a state-wide charter school system, Black/African American students in the charter school intervention demonstrated lower gains in reading than White students in public schools, but Hispanic/Latinx students receiving the intervention demonstrated larger gains in reading than White students not receiving the intervention (Center for Research on Education Outcomes, 2011).

3.4.2 Meta-analyses. Table 3 depicts the associations between meta-analysis characteristics and whether the study examined impacts on disparities. Only 13 (6%) out of 210 meta-analyses conducted analyses to examine potential effects for reducing disparities. Meta-analyses that conducted these analyses, as compared to those that did not, were more likely to test SES as a moderator, and to be published as organizational reports.

Of the 13 studies that did conduct subgroup analyses, eight found no differences in effects between racial groups (Cole, 2013, 2014; Durlak et al., 2011; Hill & Tyson, 2009; Jeynes, 2007, 2008; Li & Suen, 2012; Wilson et al., 2011). Four studies found stronger effects for non-White students; these included two meta-analyses examining language and literacy pedagogical approaches (Darrow, 2009; Gee, 1995), one examining STEM approaches for university students (Springer et al., 1999), and one examining behavioral intervention strategies for students with emotional or behavioral disorders. Finally, one meta-analysis of family-based literacy interventions found stronger effect sizes for White participants (Manz et al., 2010).

4. Discussion

This review examined the extent to which current educational intervention research a) provides complete reporting on race/ethnicity, b) includes representative samples, and c) examines disparity-reduction effects by disaggregating study results, including race/ethnicity as a moderator, or comparing effect sizes to known benchmarks. The majority of empirical studies, but not meta-analyses, provided information on the racial/ethnic composition of their samples. Among empirical studies and meta-analyses that reported race/ethnicity, White and Black/African American participants were included in most studies, and Black/African American participants were over-represented compared to their population within US public schools. Few empirical studies and meta-analyses conducted analyses to investigate implications for reducing educational disparities. When these analyses were conducted, they yielded mixed results, indicating few interventions that demonstrate promise for mitigating inequities.

4.1 Consistency and Quality of Reporting of Race/Ethnicity

The majority of empirical studies (83%) reported complete or partial racial/ethnic characteristics of their sample. Reporting practices may be driven by a consensus among federal

agencies and professional organizations that fund and publish research regarding the importance of collecting and reporting race and ethnicity data for enhancing scientific knowledge. Consistent with such governmental reporting guidelines, we found that 100% of governmental reports included race/ethnicity information. Reporting the racial/ethnic composition of a study sample allows for an understanding of the degree to which study results may generalize to the broader population. The American Psychological Association (2019) style guide recommends reporting demographic characteristics that are relevant to the study findings; the fact that more recent studies were more likely to report the racial/ethnic demographics of their samples may indicate that researchers increasingly recognize the relevance of race/ethnicity in educational outcomes. Meta-analyses, however, are not following this trend, with remarkably few of these studies (8%) including complete or partial sample demographics and no significant improvement over time.

There may be a number of reasons for inadequate and poor-quality reporting, especially amongst meta-analyses. Despite the growing consensus that reporting race/ethnicity is important, there are no clear universal standards for doing so. While there is some consistency among federal agencies regarding a common approach to collecting and reporting race and ethnicity data, these reporting methods are not applied universally, particularly when a project is not bound by the requirements of funders. In addition, the standards that are set by federal agencies may not always meet the needs or preferences of researchers. For example, scholars have advocated for research to better reflect the heterogeneity within racial/ethnic groups through the examination of more nuanced demographic categories (e.g., country/region of origin, cultural background, skin tone, generation status; Nguyen et al., 2017); the common categories suggested by federal agencies may not reflect individuals' own preferences or self-identity (e.g., Moreno & Gaytán, 2013) and may mask disparities in educational and other outcomes within a population

(e.g., Anyon et al., 2014; Manuel et al., 2012; O'Connor et al., 2007). Thus, importantly, researchers may opt to collect and report demographic data that better reflects the specific characteristics of their research population and/or that best facilitates analysis of their key research questions. Although empirical studies are able to provide complete reporting for their individual study regardless of their demographic data collection methods, the lack of standardization can raise challenges for meta-analyses; meta-analytic coding is often forced to provide the "lowest common denominator", which may result in not presenting racial/ethnic categories when individual studies use idiosyncratic approaches. While it is critical to be responsive to participant preferences and the analytic needs of specific studies, it is also beneficial for data to be reported in a way that allows for studies to build on each other. This can be facilitated if individual study authors provide complete demographic information with as much detail as possible, regardless of the approach they take, so that information can be summarized and synthesized across different collection methods.

4.2 Representation of Racial/Ethnic Groups

When considering empirical studies and meta-analyses that did report information about racial and ethnic minority youth in their research, Black/African American students are over-represented compared to the rates of public-school attending youth, while White, Hispanic/Latinx, Asian/Pacific Islander, and Native American youth are under-represented compared to the population. Students classified as Other races/ethnicities were over-represented in empirical studies, but under-represented in meta-analyses. In other fields, such as psychology, research has also found that Hispanic/Latinx, Asian, Pacific Islander, and Native American individuals are under-represented in intervention research (Huey & Polo, 2008). However, this

prior research indicates that Black/African American individuals are also under-represented and that White individuals are over-represented, which is not consistent with our review.

The higher levels of representation of Black/African American individuals in educational research, not seen in other fields, may be related to the extensive history of scholarship focused on characterizing and improving the experiences of Black/African American youth within US schools over many decades (Ladson-Billings, 2005). In fact, intervention research conducted initially with primarily Black/African American students has become the cornerstone for widespread intervention and policy-making across contexts, such as initiatives in early childhood education (Schweinhart, 2000). However, it is critical to recognize that although Black/African American students may be over-represented in educational research relative to their population in US public schools, this representation does not necessitate equitable inclusion. Educational research has often problematized Black/African American youth, communities, and families, as opposed to recognizing and building off their strengths and/or recognizing the structural biases that contribute to the disparities they face (Ladson-Billings, 2005). Many scholars have advocated for and developed frameworks outlining strengths-based approaches (e.g., Nicolas et al., 2008), and it is crucial for educational interventions to integrate such approaches (e.g., Okeke-Adeyanju et al., 2014). The under-representation of White students compared to the public-school population may also be related to the nature of educational intervention research. As discussed earlier, school-based research may be better suited than other social science fields to recruit non-White samples, and education interventions may intentionally attempt to focus on improving outcomes for students who experience educational disparities.

Our results also demonstrate that all racial/ethnic groups are not equally included in education intervention studies. Although White and Black/African American students were

included to some extent in at least 80% of empirical studies of educational interventions, the rate of inclusion was much lower for other groups. Notably, Native American students were only included in 16% of empirical studies that reported race/ethnicity and had no representation at all in any of the meta-analyses that reported race/ethnicity. Barriers to research participation for marginalized groups, such as mistrust, lack of access to information, stigma, logistical challenges, and a lack of cultural responsiveness, have been well-documented (George et al., 2014), and these may be particularly profound for indigenous populations. The principles, goals, and implications of scientific research can be considered reductionist and maintain a legacy of colonialism, and may be misaligned with the cultural values of Native American communities (Glover et al., 2015). Researchers often do not integrate the worldview or cultural wealth of knowledge of these communities into their research as a means to enhance inclusivity. In addition, due to their status as sovereign nations, research conducted within Native communities (e.g., in tribal schools) must be approved by tribal governments, in addition to university ethics review boards, thus creating additional barriers to inclusion in research, and requiring researchers to build strong, community-oriented relationships and partnerships (Harding et al., 2012). In order to enhance the inclusion of under-represented groups in educational research, serious consideration must be given to how the research question, study design, and intervention implementation may serve as barriers or facilitators for inclusive and equitable participation.

It is important to note that these results do not shed light on sample heterogeneity or homogeneity within a specific study. When research is conducted in schools, participant samples reflect the racial/ethnic breakdown of enrollment, and schools are largely segregated along racial lines (McFarland et al., 2017; Wagner, 2017). Thus, when researchers intend to recruit diverse samples reflective of the public-school population, they must follow intentional recruitment

strategies. Resources like www.theGeneralizer.org (Hedges et al., n.d.) can assist researchers with defining an inference population of public schools using criteria from publicly available data and developing a plan to recruit a sample that is similar to the inference population. The Generalizer also provides school contact information and recruitment goals to guide the process of recruitment. Including a heterogenous sample is not always the goal of a particular research study, however. There is great value in single group research to better understand intervention effects within that group (Moreno & Gaytán, 2013) or to examine culturally-specific interventions that likely engage in a strengths-based approach and make a substantial contribution to reducing disparities (G. C. N. Hall et al., 2016). Thus, we do not recommend a sole focus on recruiting heterogeneous samples for educational intervention research. Adequate representation of youth of color in research can be accomplished in multiple ways and reducing racial and ethnic disparities is a complex issue that can benefit from multiple strategies. Research designed to intentionally recruit diverse samples reflective of the overall population and research focusing on the effects of interventions within marginalized groups both contribute unique valuable information.

4.3 Effects on Reducing Disparities

Current research also provides very little information regarding educational interventions that demonstrate promise for enhancing educational equity. Only 19% of empirical studies and 6% of meta-analyses conducted moderation, differential, or benchmarking analyses to understand the effects of interventions on reducing disparities. This is a significant concern. Our review suggests that even for educational decision makers who seek out evidence-based practices to implement in their schools from resources such as WWC, there is little guidance available regarding the extent to which these programs or practices will impact pervasive academic gaps

and racial/ethnic disproportionality.

There are likely a number of reasons that researchers do not conduct or report moderation analyses or other techniques to examine the effects of interventions on educational disparities. For one, researchers may be concerned that their studies are underpowered. Reliable and valid analyses require samples large enough to detect an effect; after stratifying the sample, there may be not enough participants in subgroups to permit testing. Although there would still be benefits to reporting the results (e.g., use in future meta-analyses), it is not yet standard practice to report the results of underpowered analyses that may appear irrelevant to the researchers' primary aim. There is also publication bias (i.e. the "file drawer" problem), which pertains to the lack of motivation for researchers to write papers with non-significant or negative findings, for journals to publish these types of papers, and for the scientific community to read these types of papers (Song et al., 2010). If interventions favor White students and increase racial/ethnic disparities, researchers may be hesitant to report such findings, which may draw attention away from the overall effect of their novel program. Finally, the importance of using data disaggregated by race/ethnicity to guide decision-making and monitor progress has gained national attention only recently (Annie E. Casey Foundation, 2016). There now exist models in other fields (e.g., mental health) that can guide intervention research aimed specifically at issues of racial equity (Huey & Polo, 2008).

Of the studies that do examine the impact on disparities, most find equivalent effects across racial/ethnic groups, suggesting that these interventions are likely to maintain the disparities that already exist. Only 5% of empirical studies and 1.5% of meta-analyses demonstrated the potential for educational interventions to reduce academic and socio-emotional disparities. These studies spanned multiple content areas across Pre-K through post-secondary

contexts. Thus, although the results of these individual programs are promising, it is difficult to draw broad conclusions regarding best practices to reduce disparities. It is interesting to note that although relatively few studies (8%) included post-secondary samples overall, suggesting a relatively low number of post-secondary interventions within the WWC, post-secondary studies were over-represented (11%) among empirical studies that examined disparity-reducing effects. Of just eight empirical studies that were conducted specifically in post-secondary settings, two (25%) disaggregated intervention effects according to race/ethnicity. In addition, a third study focused on providing job skills training for youth who dropped out of school also disaggregated their findings. Due to the proximity to the transition to adulthood, and the long-term associations between education disparities and inequities in employment, earnings, and health, it is important for future research to better understand the specific types and levels of postsecondary and vocational experiences that can increase equity in both educational and societal outcomes.

The goal of enhancing equity throughout the intervention development, implementation, and evaluation processes is necessary in order to have substantial impacts on reducing disparities (Williams & Deutsch, 2016). However, examining the potential for an intervention to reduce disparities is often secondary to the primary aim of testing overall effectiveness. In contrast, there is a vast literature describing culturally responsive practices in education, which explicitly aim to reduce disparities by enhancing the fit between students' home and school cultures and addressing structural inequities within schools (Gay, 2010). Culturally responsive practices and programs, however, have rarely been evaluated using rigorous methodologies to test their effectiveness. Extensive qualitative studies provide rich descriptions of teacher practices that demonstrate the importance of cultural responsiveness (e.g., Brown, 2004; Kraft, 2007; Moll, Amanti, Neff, & Gonzalez, 1992; Monroe & Obidah, 2004), but do not provide evidence of

effectiveness that would allow for inclusion in a compendium such as WWC. A systematic review of in-service teacher interventions focused on enhancing culturally responsive practices indicated that none of these interventions met standards of evidence according to the WWC (Bottiani et al., 2018). Thus, there is a need to apply more rigorous evaluation techniques to interventions that incorporate an explicit focus on enhancing cultural responsiveness.

4.4 Limitations & Future Directions

Although this study provides novel information regarding the current state of educational intervention research with regard to race and ethnicity, it is not without limitations. First, we constrained our review to empirical studies and meta-analyses that were classified within the IES WWC. This permitted a focus on interventions that have been systematically vetted by a federal agency and are accessible by educators who are interested in adopting evidence-based practices. On the other hand, it also meant that our review was limited to certain topic areas. In addition, we only reviewed studies that IES classified as "meeting design standards," which allowed us to ensure a certain level of rigor. However, this limited our review to exclude any promising practices that have yet to meet these standards. For example, as mentioned above, research on culturally responsive practices likely include non-White samples and may demonstrate profound impacts for reducing disparities, but often does not reach a level of evidence that would warrant inclusion in this review. In addition, the IES WWC is calibrated to the priorities of the US DOE, perhaps limiting the relevance of these findings to other countries, particularly in the categorization of specific racial/ethnic groups, as well as the topic areas that are reviewed. Thus, additional research may be necessary in order to conduct similar reviews that can examine the potential for educational interventions to reduce disparities in other contexts.

Additionally, scholars have expressed important concerns regarding the use of

race/ethnicity in research. Historically, when differences in educational and other outcomes have been identified according to race or ethnicity, there has been a tendency for researchers to interpret the results from a deficit perspective, whereby disparities are attributed to deficiencies in the students (e.g., personal attributes), their families (e.g., parenting styles), communities (e.g., cultural norms and expectations; Cabrera, 2013). Racial/ethnic differences have also been examined or interpreted through the lens of genetic differences, reifying the problematic conceptualization of race as a biological construct (Boyd et al., 2020). Instead, it is essential to recognize the sociocultural context and systematic biases in educational and other social systems that contribute to such disparities (e.g., implicit and explicit biases, exclusionary discipline policies, biased curriculum, funding). A concern of the current study is that, considering the overreliance on deficit models to characterize racial/ethnic differences, disaggregating intervention results by race/ethnicity may lead to the conclusion that one's race or ethnicity is a risk factor. These analyses should instead shed light on the extent to which an intervention is designed and implemented in such a way that it can respond to the backgrounds and identities of students of color. In addition, instead of solely including racial or ethnic grouping variables in research, scholars have advocated for the inclusion of variables that represent relevant social mechanisms that may underlie racial or ethnic differences (Cheng et al., 2015; Helms et al., 2005). This would provide information regarding how interventions can more effectively and equitably address malleable factors that contribute to racial/ethnic disparities. It is also important to include variables that allow for the possibility of disentangling the influence of various social indicators (e.g., SES, language, gender) from race/ethnicity. In the current review, as well as in many of the included studies, race/ethnicity was included as a static grouping variable. We recognize the significant limitations of this approach and urge more education researchers to

consider the underlying societal and structural mechanisms that may impact intervention implementation and outcomes with diverse groups.

4.5 Implications

The implications of this paper extend to researchers, practitioners, and policy makers. This paper serves as a call for researchers to be more intentional in designing and evaluating educational interventions that can better address the needs of marginalized students and reduce pervasive disparities in educational outcomes. First, intentional recruitment strategies are essential to ensure that research samples represent the intended population. As we argue above, strategic samples may mean recruiting a heterogenous sample with potential for broad generalization, or a relatively homogenous sample that allows for in-depth exploration of social mechanisms that are relevant to that specific population. In addition, there continues to be a need for clear reporting of the racial/ethnic composition of samples. Standards for reporting on race and ethnicity can continue to be refined by professional institutions, governmental organizations, and journal leadership. Finally, the need for more systematic assessment of the impact of interventions on educational disparities is evident. Funders and publishers can increase the expectations that researchers conduct and report subgroup or other relevant single group analyses, regardless of null or contradictory findings. Additionally, researchers can contribute to future meta-analyses by making data available via the multitude of "Open Education Science" practices and resources that are now emerging (van der Zee & Reich, 2018). This would greatly permit sufficiently powered meta-analyses to capitalize on existing underpowered samples. IES WWC could include a designation for studies that a) tested the effects of an intervention for reducing disparities and b) demonstrated significant disparity-reducing effects. Clear reporting ensures this essential information is easily accessible for practitioners as they make choices

regarding programs that would be most suitable for their context. It is important for educational decision makers at all levels to utilize the race/ethnicity information that is reported to ensure that the interventions they implement are appropriate for their student population.

4.6 Conclusion

This systematic review examined the extent to which current educational intervention research reports on the racial/ethnic characteristics of their samples, includes representative participant samples, and evaluates the implications of interventions for reducing pervasive educational disparities. Our review demonstrated that many empirical studies, but few meta-analyses, report on race/ethnic sample characteristics and that many students of color are often under-represented across the body of educational research. Our review also revealed that very few studies provide adequate information regarding the potential for educational interventions to reduce disparities. These findings bring to light the pressing need for educational interventions to prioritize research questions related to race, ethnicity, and equity, across educational domains and levels.

References

Alibali, M. W., & Nathan, M. J. (2010). Conducting research in schools: A practical guide. *Journal of Cognition and Development*, *11*(4), 397–407. https://doi.org/10.1080/15248372.2010.516417

American Psychological Association. (2019). *Publication Manual of the American Psychological Association* (7th Edition). American Psychological Association.

Annie E. Casey Foundation. (2016, August 8). *Taking Data Apart: Why a Data-Driven Approach Matters to Race Equity*. The Annie E. Casey Foundation. https://www.aecf.org/blog/taking-data-apart-why-a-data-driven-approach-matters-to-race-equity/

Anyon, Y., Ong, S. L., & Whitaker, K. (2014). School-based mental health prevention for Asian American adolescents: Risk behaviors, protective factors, and service use. *Asian American Journal of Psychology*, *5*(2), 134–144. https://doi.org/10.1037/a0035300

Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *The American Psychologist*, *63*(7), 602–614. https://doi.org/10.1037/0003-066X.63.7.602

Auguste, B., Hancock, B., & Laboissiere, M. (2009). *The economic cost of the US education gap: McKinsey*. McKinsey & Company. https://www.mckinsey.com/industries/social-sector/our-insights/the-economic-cost-of-the-us-education-gap

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series*

B (Methodological), *57*(1), 289–300.

Blom-Hoffman, J., Leff, S. S., Franko, D. L., Weinstein, E., Beakley, K., & Power, T. J. (2009). Consent procedures and participation rates in school-based intervention and prevention research: Using a multi-component, partnership-based approach to recruit participants. *School Mental Health*, *1*(1), 3–15. https://doi.org/10.1007/s12310-008-9000-7

Bottiani, J. H., Larson, K. E., Debnam, K. J., Bischoff, C. M., & Bradshaw, C. P. (2018). Promoting educators' use of culturally responsive practices: A systematic review of inservice interventions. *Journal of Teacher Education*, *69*(4). https://doi.org/10.1177/0022487117722553

Boyd, R. W., Lindo, E. G., Weeks, L. D., & McLemore, M. R. (2020). On racism: A new standard for publishing on racial health inequities | health affairs. *Health Affairs Blog*. https://www.healthaffairs.org/do/10.1377/hblog20200630.939347/full/?utm_source=Newsletter&utm_medium=email&utm_content=Racism+++COVID-19%3B+Access+To+ACA+Coverage+In+The+COVID-19+Crisis%3B+Nonphysician+Practitioners+In+ACOs&utm_campaign=HASU+7-5-20

Bradshaw, C. P., Pas, E. T., Bottiani, J. H., Debnam, K. J., Reinke, W. M., Herman, K. C., & Rosenberg, M. S. (2018). Promoting cultural responsivity and student engagement through Double Check coaching of classroom teachers: An efficacy study. *School Psychology Review*, *47*(2), 118–134.

Brotman, L. M., Dawson-McClure, S., Calzada, E. J., Huang, K.-Y., Kamboukos, D., Palamar, J. J., & Petkova, E. (2013). Cluster (school) RCTof ParentCorps: Impact on kindergarten

academic achievement. *Pediatrics*, *131*(5), e1521–e1529. https://doi.org/10.1542/peds.2012-2632

Brown, D. F. (2004). Urban teachers' professed classroom management strategies: Reflections of culturally responsive teaching. *Urban Education*, *39*(3), 266–289. https://doi.org/10.1177/0042085904263258

Cabrera, N. J. (2013). Positive development of minority children. *SRCD Social Policy Report*, *27*(2), 3–22.

Caspi, A., Entner Wright, B. R., Moffitt, T. E., & Silva, P. A. (1998). Early failure in the labor market: Childhood and adolescent predictors of unemployment in the transition to adulthood. *American Sociological Review; Washington*, *63*(3), 424–451.

Castro-Olivo, S. M. (2017). Introduction to special issue: Culturally responsive school-based mental health interventions. *Contemporary School Psychology*, *21*(3), 177–180. https://doi.org/10.1007/s40688-017-0137-y

Cave, G., Bos, H., Doolittle, F., & Toussaint, C. (1993). *JOBSTART: Final report on a program for school dropouts.* MDRC.

Center for Research on Education Outcomes. (2011). *Charter school performance in Indiana*. Author.

Cheng, T. L., Goodman, E., & The Committee on Pediatric Research. (2015). Race, ethnicity, and socioeconomic status in research on child health. *PEDIATRICS*, *135*(1), e225–e237. https://doi.org/10.1542/peds.2014-3109

Cole, M. W. (2013). Rompiendo el silencio: Meta-analysis of the effectiveness of peer-mediated learning at improving language outcomes for ELLs. *Bilingual Research Journal*, *36*(2), 146–166. https://doi.org/10.1080/15235882.2013.814609

Cole, M. W. (2014). Speaking to read: Meta-analysis of peer-mediated learning for English Language Learners. *Journal of Literacy Research*, *46*(3), 358–382. https://doi.org/10.1177/1086296X14552179

Darrow, C. L. (2009). *Language and literacy effects of curriculum interventions for preschools serving economically disadvantaged children: A meta analysis*. Society for Research on Educational Effectiveness, Crystal City, VA.

Dobbie, W., & Fryer, R. (2009). *Are high quality schools enough to close the achievement gap? Evidence from a social experiment in Harlem* (No. w15473). National Bureau of Economic Research. https://doi.org/10.3386/w15473

Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Weissberg, R. P., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405–432. JSTOR.

Esbensen, F.-A., Miller, M. H., Taylor, T., He, N., & Freng, A. (1999). Differential Attrition Rates and Active Parental Consent. *Evaluation Review*, *23*(3), 316–335. https://doi.org/10.1177/0193841X9902300304

FDA Office of Minority Health. (2016). *Collection of race and ethnicity data in clinical trials*. U.S. Department of Health and Human Services.

Gabler, N. B., Duan, N., Liao, D., Elmore, J. G., Ganiats, T. G., & Kravitz, R. L. (2009). Dealing with heterogeneity of treatment effects: Is the literature up to the challenge? *Trials*, *10*(1), 43. https://doi.org/10.1186/1745-6215-10-43

Gay, G. (2010). *Culturally Responsive Teaching: Theory, Research, and Practice* (2nd ed.). Teachers College Press.

Gee, E. J. (1995). *The effects of a whole language approach to reading instruction on reading*

comprehension: A meta-analysis. American Educational Research Association, San Francisco, CA. https://files.eric.ed.gov/fulltext/ED384003.pdf

George, S., Duran, N., & Norris, K. (2014). A Systematic Review of Barriers and Facilitators to Minority Research Participation Among African Americans, Latinos, Asian Americans, and Pacific Islanders. *American Journal of Public Health*, *104*(2), e16–e31. https://doi.org/10.2105/AJPH.2013.301706

Gleason, P., Clark, M., Tuttle, C. C., & Dwoyer, E. (2010). *The evaluation of charter school impacts: Final report (NCEE 2010-4029).* National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Glover, M., Kira, A., Johnston, V., Walker, N., Thomas, D., Chang, A. B., Bullen, C., Segan, C. J., & Brown, N. (2015). A systematic review of barriers and facilitators to participation in randomized controlled trials by Indigenous people from New Zealand, Australia, Canada and the United States. *Global Health Promotion*, *22*(1), 21–31. https://doi.org/10.1177/1757975914528961

Green, L. R., Blasik, K., Hartshorn, K., & Shatten-Jones, E. (2000). Closing the achievement gap in science: A program to encourage minority and female students to participate and succeed. *ERS Spectrum*, *18*(2), 3–13.

Gregory, A., Huang, F. L., Anyon, Y., Greer, E., Downing, B., & Bradshaw, C. (2018). An examination of restorative interventions and racial equity in out-of-school suspensions. *School Psychology Review*, *47*(2), 167–182. https://doi.org/10.17105/SPR-2017-0073.V47-2

Gunn, B., Biglan, A., Smolkowksi, K., & Ary, D. (2000). The efficacy of supplemental

instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. *Journal of Special Education*, *34*(2), 90–103.

Hall, B. W., & Bacon, T. P. (2006). Building a foundation against violence: Impact of a school-based prevention program on elementary students. *Journal of School Violence*, *4*(4), 63–83. https://doi.org/10.1300/J202v04n04_05

Hall, G. C. N., Ibaraki, A. Y., Huang, E. R., Marti, C. N., & Stice, E. (2016). A meta-analysis of cultural adaptations of psychological interventions. *Behavior Therapy*, *47*(6), 993–1014. https://doi.org/10.1016/j.beth.2016.09.005

Harackiewicz, J. M., Canning, E. A., Tibbetts, Y., Giffen, C. J., Blair, S. S., Rouse, D. I., & Hyde, J. S. (2014). Closing the social class achievement gap for first-generation students in undergraduate biology. *Journal of Educational Psychology*, *106*(2), 375–389. https://doi.org/10.1037/a0034679

Harding, A., Harper, B., Stone, D., O'Neill, C., Berger, P., Harris, S., & Donatuto, J. (2012). Conducting research with tribal communities: Sovereignty, ethics, and data-sharing issues. *Environmental Health Perspectives*, *120*(1), 6–10. https://doi.org/10.1289/ehp.1103904

Hedges, L. V., Tipton, E., & Miller, K. (n.d.). *Generalizer*. The Generalizer. Retrieved April 26, 2019, from https://www.thegeneralizer.org/

Helms, J. E., Jernigan, M., & Mascher, J. (2005). The meaning of race in psychology and how to change it: A methodological perspective. *American Psychologist*, *60*(1), 27–36.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*(2–3), 61–83; discussion 83-135. https://doi.org/10.1017/S0140525X0999152X

Hill, N. E., & Tyson, D. F. (2009). Parental involvement in middle school: A meta-analytic assessment of the strategies that promote achievement. *Developmental Psychology*, *45*(3), 740–763. https://doi.org/10.1037/a0015362

Huang, G., Reiser, M., Parker, A., Muniec, J., & Salvucci, S. (2003). *Institute of Education Sciences findings from interviews with education policy-makers*.

Huey, S. J., & Polo, A. J. (2008). Evidence-based psychosocial treatments for ethnic minority youth: *Journal of Clinical Child & Adolescent Psychology*, *37*(1), 262–301.

Jastrzab, J. (1997). *Youth Corps: Promising strategies for young people and their communities*. Abt Associates.

Jeynes, W. H. (2007). The relationship between parental involvement and urban secondary school student academic achievement: A meta-analysis. *Urban Education*, *42*(1), 82–110. https://doi.org/10.1177/0042085906293818

Jeynes, W. H. (2008). A meta-analysis of the relationship between phonics instruction and minority elementary school student academic achievement. *Education and Urban Society*, *40*(2), 151–166. https://doi.org/10.1177/0013124507304128

Kraft, M. (2007). Toward a school-wide model of teaching for social justice: An examination of the best practices of two small public schools. *Equity & Excellence in Education*, *40*(1), 77–86. https://doi.org/10.1080/10665680601076601

Ladson-Billings, G. (2005). Foreward. In J. E. King (Ed.), *Black Education: A Transformative Research and Action Agenda for the New Century* (pp. xiii–xvii). American Education Research Association.

Li, H., & Suen, H. K. (2012). The effects of test accommodations for English language learners: A meta-analysis. *Applied Measurement in Education*, *25*(4), 327–346.

https://doi.org/10.1080/08957347.2012.714690

Lochman, J. E., & Wells, K. C. (2004). The Coping Power program for preadolescent aggressive boys and their parents: Outcome effects at the 1-year follow-up. *Journal of Consulting and Clinical Psychology*, *72*(4), 571–578. https://doi.org/10.1037/0022-006X.72.4.571

Manuel, R. C., Taylor, R. J., & Jackson, J. S. (2012). Race and ethnic group differences in socio-economic status: Black Caribbeans, African Americans, and non-Hispanic Whites in the United States. *Western Journal of Black Studies*, *36*(3), 228–239.

Manz, P. H., Hughes, C., Barnabas, E., Bracaliello, C., & Ginsburg-Block, M. (2010). A descriptive review and meta-analysis of family-based emergent literacy interventions: To what extent is the research applicable to low-income, ethnic-minority or linguistically-diverse young children? *Early Childhood Research Quarterly*, *25*(4), 409–431. https://doi.org/10.1016/j.ecresq.2010.03.002

McFarland, J., Hussar, B., DeBrey, C., Snyder, T., Wang, X., Wilkinson-Flicker, S., Gebrekristos, S., Zhang, J., Rathbun, A., Barmer, A., Mann, F. B., & Hinz, S. (2017). *The Condition of Education 2017* (No. 2017–144; NCES, p. 386). US Department of Education.

Moll, L. C., Amanti, C., Neff, D., & Gonzalez, N. (1992). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory into Practice*, *31*(2), 132–141.

Monroe, C. R., & Obidah, J. E. (2004). The influence of cultural synchronization on a teacher's perceptions of disruption: A case study of an African American middle-school classroom. *Journal of Teacher Education*, *55*(3), 256–268.

https://doi.org/10.1177/0022487104263977

Moreno, G., & Gaytán, F. X. (2013). Focus on Latino learners: Developing a foundational understanding of Latino cultures to cultivate student success. *Preventing School Failure: Alternative Education for Children and Youth*, *57*(1), 7–16. https://doi.org/10.1080/1045988X.2013.731271

National Center for Education Statistics. (2014). *National Assessment of Education Progress Data Explorer*.

National Center for Education Statistics. (2017). *Enrollment and percentage distribution of enrollment in public elementary and secondary schools, by race/ethnicity and region: Selected years, fall 1995 through fall 2027*. https://nces.ed.gov/programs/digest/d17/tables/dt17_203.50.asp?current=yes

National Joint Committee on Learning Disabilities. (1991). *The need for subject descriptors in learning disabilities research: Preschool through high school years*. American Speech-Language-Hearing Associatio. https://doi.org/10.1044/policy.PS1991-00099

Nguyen, B. M. D., Hernandez, E., Freeman, R., Alcantar, C. M., Kim, V., Nguyen, M. H., Curammeng, E. R., Paredes, A. D., & Teranishi, R. T. (2017). *The racial heterogeneity project: Implications for educational research, practice, and policy*. Institute for Immigration, Globalization, and Education & ACT Center for Equity in Learning.

Nicolas, G., Helms, J. E., Jernigan, M. M., Sass, T., Skrzypek, A., & DeSilva, A. M. (2008). A conceptual framework for understanding the strengths of Black youths. *Journal of Black Psychology*, *34*(3), 261–280. https://doi.org/10.1177/0095798408316794

O'Connor, C., Lewis, A., & Mueller, J. (2007). Researching "Black" educational experiences

and outcomes: Theoretical and methodological considerations. *Educational Researcher*, *36*(9), 541–552. https://doi.org/10.3102/0013189X07312661

Office of Management and Budget. (1997). *Revisions to the standards for the classification of federal data on race and ethnicity* (Notice No. 97–28653; pp. 58782–58790). Office of Management and Budget (OMB). https://www.whitehouse.gov/wp-content/uploads/2017/11/Revisions-to-the-Standards-for-the-Classification-of-Federal-Data-on-Race-and-Ethnicity-October30-1997.pdf

Office of Minority Health. (2018, June 1). *Final Standards—The Office of Minority Health*. https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=3&lvlid=53

Okeke-Adeyanju, N., Taylor, L. C., Craig, A. B., Smith, R. E., Thomas, A., Boyle, A. E., & DeRosier, M. E. (2014). Celebrating the strengths of Black youth: Increasing self-esteem and implications for prevention. *The Journal of Primary Prevention*, *35*(5), 357–369. https://doi.org/10.1007/s10935-014-0356-1

Orfield, G. (2013). Housing Segregation Produces Unequal Schools: Causes and Solutions. In P. L. Carter & K. G. Welner (Eds.), *Closing the Opportunity Gap* (pp. 40–60). Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199982981.001.0001/acprof-9780199982981-chapter-4

Orfield, G., & Frankenberg, E. (2014). Increasingly segregated and unequal schools as courts reverse policy. *Educational Administration Quarterly*, *50*(5), 718–734. https://doi.org/10.1177/0013161X14548942

Overbay, A., & Baenen, N. (2003). *Fast ForWord evaluation, 2002-03 (Eye on Evaluation, E&R Report No. 03.24).* Wake County Public School System.

Pyke, C., Lynch, S., Kuipers, J., Szesze, M., & Watson, W. (2004). *Implementation study of The Real Reasons for Seasons (2003-2004): SCALE-uP Report No. 4* (No. 4; SCALE-UP). George Washington University & Montgomery County Public Schools.

Quinn, S. C., Garza, M. A., Butler, J., Fryer, C. S., Casper, E. T., Thomas, S. B., Barnard, D., & Kim, K. H. (2012). Improving informed consent with minority participants: Results from researcher and community surveys. *Journal of Empirical Research on Human Research Ethics*, *7*(5), 44–55. https://doi.org/10.1525/jer.2012.7.5.44

Reardon, S. F., Grewal, E. T., Kalogrides, D., & Greenberg, E. (2012). Brown fades: The end of court-ordered school desegregation and the resegregation of American public schools. *Journal of Policy Analysis and Management*, *31*(4), 876–904. https://doi.org/10.1002/pam.21649

Reed, D. K., McCray Sorrells, A., Cole, H. A., & Takakawa, N. N. (2013). The ecological and population validity of reading interventions for adolescents: Can effectiveness be generalized? *Learning Disability Quarterly*, *36*(3), 131–144. https://doi.org/10.1177/0731948712451976

Reingold, D. A., & Liu, H. K. (2009). Do poverty attitudes of social service agency directors influence organizational behavior? *Nonprofit and Voluntary Sector Quarterly*, *38*(2), 307–332. https://doi.org/10.1177/0899764008316967

Resendez, M., Fahmy, A., & Manley, M. A. (2005). *The relationship between using Saxon Middle School Math and student performance on Texas statewide assessments [Sample 3]*. PRES Associates, Inc.

Roosa, M. W., Liu, F. F., Torres, M., Gonzales, N. A., Knight, G. P., & Saenz, D. (2008). Sampling and recruitment in studies of cultural influences on adjustment: A case study with

Mexican Americans. *Journal of Family Psychology*, *22*(2), 293–302. https://doi.org/10.1037/0893-3200.22.2.293

Rosenberg, M. S., Bott, D., Majsterek, D., Chiang, B., Gartland, D., Wesson, C., Graham, S., Smith-Myles, B., Miller, M., & Swanson, H. L. (1992). Minimum standards for the description of participants in learning disabilities research (CLD Research Committee). *Learning Disability Quarterly*, *15*(1), 65–70.

Sarama, J., Clements, D. H., Wolfe, C. B., & Spitler, M. E. (2012). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies. *Journal of Research on Educational Effectiveness*, *5*(2), 105–135. https://doi.org/10.1080/19345747.2011.627980

Schweinhart, L. J. (2000). The High/Scope Perry preschool study: A case study in random assignment. *Evaluation & Research in Education*, *14*(3–4), 136–147. https://doi.org/10.1080/09500790008666969

Sinclair, J., Hansen, S. G., Machalicek, W., Knowles, C., Hirano, K. A., Dolata, J. K., Blakely, A. W., Seeley, J., & Murray, C. (2018). A 16-Year Review of Participant Diversity in Intervention Research Across a Selection of 12 Special Education Journals. *Exceptional Children*, *84*(3), 312–329. https://doi.org/10.1177/0014402918756989

Skiba, R. J., Arredondo, M. I., & Williams, N. T. (2014). More than a metaphor: The contribution of exclusionary discipline to a school-to-prison pipeline. *Equity & Excellence in Education*, *47*(4), 546–564. https://doi.org/10.1080/10665684.2014.958965

Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The Urban*

Review, *34*(4), 317–342.

Smith, D. D., Deshler, D., Hallahan, D., Lovitt, T., Robinson, S., Voress, J., & Ysseldyke, J. (1984). Minimum standards for the description of subjects in learning disabilities research reports. *Learning Disability Quarterly*, *7*(3), 221–225. JSTOR. https://doi.org/10.2307/1510478

Sohoni, D., & Saporito, S. (2009). Mapping school segregation: Using gis to explore racial segregation between schools and their corresponding attendance areas. *American Journal of Education*, *115*(4), 569–600. https://doi.org/10.1086/599782

Song, F., Perekh, S., Hooper, L., Loke, Y. K., Ryder, J., Sutton, A. J., Hing, C., Kwok, C. S., Pang, C., & Harvey, I. (2010). Dissemination and publication of research findings: An updated review of related biases. *Health Technology Assessment*, *14*(8). https://doi.org/10.3310/hta14080

Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, *69*(1), 21–51.

Stevens, R. J., & Durkin, S. (1992). *Using student team reading and student team writing in middle schools: Two evaluations. (Tech. Rep. No. 36).* Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students.

US Department of Education. (2017). *Digest of Education Statistics, 2017*. Institute for Education Sciences. https://nces.ed.gov/programs/digest/d17/tables/dt17_209.10.asp?current=yes

van der Zee, T., & Reich, J. (2018). Open Education Science. *AERA Open*, *4*(3), 2332858418787466. https://doi.org/10.1177/2332858418787466

Vincent, C. G., & Tobin, T. J. (2011). The relationship between implementation of School-Wide Positive Behavior Support (SWPBS) and disciplinary exclusion of students from various ethnic backgrounds with and without disabilities. *Journal of Emotional and Behavioral Disorders*, *19*(4), 217–232. https://doi.org/10.1177/1063426610377329

Wagner, C. (2017). *School Segregation Then and Now: How to Move Toward a More Perfect Union*. Center for Public Education. http://www.centerforpubliceducation.org/system/files/School%20Segregation%20Full%20Report_0.pdf

Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science*, *331*(6023), 1447–1451. https://doi.org/10.1126/science.1198364

Williams, J. L. 1, & Deutsch, N. L. 1. (2016). Beyond between-group differences: Considering race, ethnicity, and culture in research on positive youth development programs. *Applied Developmental Science*, *20*(3), 203–213. https://doi.org/10.1080/10888691.2015.1113880

Wilson, S. J., Tanner-Smith, E. E., Lipsey, M. W., Steinka-Fry, K., & Morrison, J. (2011). *Dropout prevention and intervention programs: Effects on school completion and dropout among school-aged children and youth*. The Campbell Collaboration. https://doi.org/10.4073/csr.2011.8

Witte, J. F., Cowen, J. M., Fleming, D. J., Wolf, P. J., Condon, M. R., & Lucas-McLean, J. (2010). *The MPCP Longitudinal Educational Growth Study third year report (SCDP Milwaukee Evaluation Report #15).* University of Arkansas, School Choice Demonstration

Project.

Table 1. Study descriptive statistics

	Empirical Studies		Meta-analyses	
	N	**%**	**N**	**%**
N	96	100.0	210	100.0
Grade level[1]				
Pre-K	16	16.7	49	9.9
Elementary	46	47.9	155	31.3
Middle	31	32.3	138	27.9
High	18	18.8	114	23.0
Post-secondary	8	8.3	39	7.9
Missing	3	3.1	28	13.3
Type of publication				
Academic journal	46	47.9	174	82.9
Dissertation/thesis	9	9.4	23	10.9
Government report	10	10.4	1	0.5
Organization report	31	32.3	9	4.3
Book chapter	0	0.0	3	1.4
Reported % ELL	19	19.8	NA	NA
Reported % FRL	32	33.3	NA	NA
Analyzed intervention effects for reducing disparities	18	18.8	13	6.2
Analyzed SES as a moderator	12	12.5	4	1.9
Quality of reporting on Race/ethnicity				
None provided	26	27.1	197	93.8
Partial	10	10.4	5	2.4
Complete	60	62.5	8	3.8
	M	**SD**	**M**	**SD**
Sample size	1,259	2,505	--	--
Year of publication	1989	6.9	2006	8.3

[1] Because studies could span multiple grade levels, grade level does not add to 100%

Table 2. Empirical study qualities associated with reporting of race/ethnicity, and testing for disparity reduction (n = 96)

	Reported Race/ethnicity						Stratified or moderated analyses by race/ethnicity					
	No		Yes				No		Yes			
	n	*%*	*n*	*%*	χ^2	*p*	*n*	*%*	*n*	*%*	χ^2	*p*
N/%	26	27.1%	70	72.9%			78	81.3%	18	18.8%		
Examined effectiveness for reducing disparities	**1**	**5.6%**	**17**	**94.4%**	**5.20**	**0.023**	--	--	--	--	--	--
Reported race/ethnicity	--	--	--	--	--	--	**54**	**76.1%**	**17**	**23.9%**	**5.09**	**0.024**
Sample size (quartiles)					**14.38**	**0.006**					8.24	0.083
10-64	**11**	**47.8%**	**12**	**52.2%**	**6.59**	**0.010**	21	91.3%				
65-309	6	26.1%	17	73.9%	0.02	0.902	21	91.3%				
310-1,272	4	17.4%	19	82.6%	1.44	0.230	**15**	**65.2%**				
1,273-15,661	**3**	**12.0%**	**22**	**88.0%**	**3.89**	**0.048**	20	80.0%				
Missing	**2**	**100.0%**	**0**	**0.0%**	**5.50**	**0.019**	1	50.0%				
% FRL reported	**2**	**6.3%**	**30**	**93.8%**	**10.55**	**0.001**	23	71.9%	9	28.1%	2.77	0.096
% ELL reported	2	10.5%	17	89.5%	3.29	0.070	**12**	**63.2%**	**7**	**36.8%**	**5.09**	**0.024**
% IEP reported	7	23.3%	23	76.7%	0.31	0.577	25	83.3%	5	16.7%	0.12	0.724
Grade level (category)												
Pre-K	6	37.5%	10	62.5%	1.06	0.304	14	87.5%	2	12.5%	0.49	0.483
Elementary (K-5)	14	30.4%	32	69.6%	0.50	0.478	38	82.6%	8	17.4%	0.11	0.744
Middle (6-8)	8	25.8%	23	74.2%	0.04	0.846	23	74.2%	8	25.8%	1.50	0.221
High (9-12)	5	27.8%	13	72.2%	0.01	0.941	15	83.3%	3	16.7%	0.06	0.802
Post-secondary	2	25.0%	6	75.0%	0.02	0.810	6	75.0%	2	25.0%	0.22	0.636
Missing	0	0.0%	3	100.0%	1.15	0.284	**1**	**33.3%**	**2**	**66.7%**	**4.67**	**0.031**
Publication type (category)					**11.15**	**0.011**					3.38	0.337
Academic journal	11	23.9%	35	76.1%	0.45	0.503	39	84.8%	7	15.2%		
Dissertation/Thesis	**6**	**66.7%**	**3**	**33.3%**	**7.88**	**0.005**	8	88.9%	1	11.1%		
Government report	**0**	**0.0%**	**10**	**100.0%**	**4.15**	**0.042**	9	90.0%	1	10.0%		
Organizational report	9	29.0%	22	71.0%	0.09	0.767	22	71.0%	9	29.0%		
	r(pb)	*p*					*r(pb)*	*p*				
Year of publication	**0.291**	**0.004**					0.136	0.188				
FRL percent	0.098	0.588					-0.124	0.500				

Table 3.
Meta-analysis study qualities associated with reporting of race/ethnicity, and testing for disparity reduction (n = 210)

	Reported Race/ethnicity						Stratified or moderated analyses by race/ethnicity					
	No		Yes				No		Yes			
	n	*%*	*n*	*%*	χ^2	*p*	*n*	*%*	*n*	*%*	χ^2	*p*
N/%	197	93.8%	13	6.2%			197	93.8%	13	6.2%		
Number of studies included (quartiles)					0.65	0.885					0.46	0.927
0-16	42	91.3%	4	8.7%			44	95.7%	2	4.3%		
17-28	54	94.7%	3	5.3%			53	93.0%	4	7.0%		
29-54	52	94.5%	3	5.5%			51	92.7%	4	7.3%		
55-249	49	94.2%	3	5.8%			49	94.2%	3	5.8%		
Examined effectiveness for reducing disparities	12	92.3%	1	7.7%	0.05	0.817	--	--	--	--	--	--
Tested SES as a moderator	30	100.0%	0	0.0%	2.31	0.129	**27**	90.0%	3	10.0%	.88	0.350
Grade level (category)												
Pre-K	45	91.8%	4	8.2%	0.43	0.513	44	89.8%	5	10.2%	1.77	0.183
Elementary	145	93.5%	10	6.5%	0.07	0.792	147	94.8%	8	5.2%	1.08	0.299
Middle	**126**	**91.3%**	**12**	**8.7%**	**4.35**	**0.037**	131	94.9%	7	5.1%	0.87	0.352
High	**102**	**89.5%**	**12**	**10.5%**	**8.07**	**0.004**	107	93.9%	7	6.1%	0.00	0.974
Post-secondary	37	94.9%	2	5.1%	0.09	0.760	38	97.4%	1	2.6%	1.09	0.298
Missing	28	100%	0	0.0%	2.13	.144	27	96.4%	1	3.6%	.382	.537
Publication type (category)					6.87	0.143					**13.26**	**<.01**
Academic journal	166	95.4%	8	4.6%			164	94.3%	10	5.7%	.344	0.558
Dissertation/Thesis	20	87.0%	3	13.0%			23	100.0%	0	0.0%	1.70	0.192
Government report	1	100.0%	0	0.0%			1	100.0%	0	0.0%	0.07	0.797
Organizational report	8	88.9%	1	11.1%			**6**	**66.7%**	**3**	**33.3%**	**11.93**	**<.001**
Book chapters	2	66.7%	1	33.3%			3	100.0%	0	0.0%	0.20	0.654
	r(pb)	*p*					*r(pb)*	*p*				
Year of publication	0.024	0.733					0.043	0.537				

Table 4.
Percentages of racial and ethnic minority youth represented in public schools and in educational intervention research

	U.S. Public school-attending youth (2015)	Empirical studies (of studies providing demographics)					Meta-analyses (of studies providing demographics)				
		Group mean %[1]				% studies including group[2]	Group mean %[1]				% studies including group[2]
	%	%	SD	SE	OR[3]	%	%	SD	SE	OR	%
White	49	39.9	24.8	3.3	0.69	82.9	41.9	25.3	7.0	0.75	84.6
Black/African American	15	34.9	25.6	3.4	3.04	80.0	34.5	30.5	8.5	2.98	84.6
Hispanic/Latinx	26	13.8	14.9	2.0	0.46	69.5	7.7	12.1	3.4	0.23	53.8
Asian/Pacific Islander	5	2.9	7.1	0.9	0.57	32.9	0.5	1.3	0.4	0.10	15.4
Native American	1	0.6	2.3	0.3	0.60	15.7	0.0	0.0	0.0	--	0.0
Other	3	6.4	14.5	1.9	2.21	44.3	2.1	4.2	1.2	0.69	23.1
Missing[4]	--	--	--	--		--	13.9	15.0	4.2	--	61.5

[1] The group mean percentage is the mean of each study's percentages
[2] This percentage is the number of studies that include each group out of the total number of studies that did report race/ethnicity (empirical studies = 71, meta-analyses = 13)
[3]OR = Odds Ratio, OR = $(p_2 / 1 - p_2) / (p_1 / 1 - p_1)$, where p_2 represents the proportion of a particular racial/ethnic group within empirical studies/meta-analyses and p_1 represents the proportion of a particular racial/ethnic group within the population of public-school attending youth
[4] "Missing" in meta-analyses resulted from lack of demographics reporting from primary studies within the meta-analysis

Table 5.
Detailed descriptions of empirical studies and meta-analyses that examine intervention effects on racial/ethnic disparities

Citation	Research Design; Intervention Focus	Analytic Approach	Results of Analyses
Empirical Studies			
Brotman et al., 2013	Randomized Control Trial; School Readiness	Compared program effect sizes with established performance gaps	**Effects approach magnitude of disparity:** Program effect sizes approach magnitude of achievement gaps, and match or exceed effect sizes of other school-based interventions
Cave et al., 1993	Randomized Control Trial; Dropout prevention	F-test examining interaction between experimental condition and racial category	**Equivalent results across groups:** Positive impacts of JOBSTART on program outcomes (e.g., receipt of GED, earnings) did not differ according to race/ethnicity
Center for Research on Education Outcomes, 2011	Quasi-Experimental; Comprehensive School Reform	Compared test scores of students in charter schools with a matched pair in public school	**Mixed findings:** Black students in charter schools demonstrated lower levels of reading achievement, but no significant differences in math achievement, than White students in public schools. Hispanic students in charter schools demonstrated higher levels of reading achievement, but no significant differences in math achievement, than White students in public schools.
Dobbie & Fryer, 2009	Randomized Lottery; Comprehensive School Reform	Compared test scores of students in the experimental school to a matched sample not randomized into the school, and average test scores for White and Black NYC public school students	**Effects approach magnitude of disparity:** Significant disparities in ELA and Math scores between program students and average White public-school students before school entry. Disparities eliminated in ELA and Math by end of elementary; disparities eliminated in Math and reduced in ELA by end of middle school
Gleason et al., 2010	Randomized Lottery; Comprehensive School Reform	Compared test scores of students who "won" lottery to attend charter program to those who "lost" lottery	**Equivalent results across groups:** Charter schools did not impact academic outcomes (reading or math) for any students, including when broken down by race
Gunn et al., 2000	Randomized Control Trial; Literacy	F-text examining interaction between experimental condition and ethnicity	**Stronger effects for non-Hispanic students:** Non-Hispanic children had greater gains in program outcomes (e.g., vocabulary, oral reading fluency) than Hispanic children

Hall & Bacon, 2005	Randomized Control Trial; Student Behavior	Described disaggregated findings by race and gender	**Equivalent results across groups:** White, African American, and Hispanic students in the intervention condition demonstrated significant improvements in program outcomes (i.e., emotional competency, social and resistance skills, communication, interactions with others)
Harackiewicz et al., 2014	Randomized Control Trial; Postsecondary Education	Disaggregated study results according to race/ethnicity and generational status	**Stronger effects for White students:** Intervention effects (class performance, GPA, continuation) were not significant for under-represented minorities (African American, Hispanic, or Native Americans), but positive effects were found for majority students
Jastrzab, et al., 1997	Randomized Control Trial; Dropout Prevention	Described disaggregated findings by race and gender	**Stronger effects for non-White students:** Program completion rates highest for Latino and Asian American participants; African American and Hispanic men and women and White women demonstrated the most significant, positive program outcomes (e.g., employment, civic engagement, educational attainment)
Lochman et al., 2004	Randomized Control Trial; Student Behavior	F-text examining interaction between experimental condition and ethnicity (White vs. non-White)	**Mixed findings:** Intervention reduced substance use and improved school behavior for White boys, but not non-White boys; Intervention reduced delinquency equally for White and non-White boys
Overbay & Baenen, 2003	Quasi-Experimental; Literacy	Predicted program changes by race	**Mixed findings:** Reading improved for all students in grades 3-8. For sixth graders, Non-White students demonstrated significantly lower growth than White students.
Pyke et al., 2004	Randomized Control Trial; Dropout Prevention	F-test examining interaction between experimental condition and race/ethnicity categories	**Equivalent results across groups:** No race/ethnicity x condition interactions for any program outcomes
Resendez et al., 2005	Quasi-experimental; Mathematics	Disaggregated growth rate over time for White, Hispanic, and African American students within intervention group	**Stronger effects for non-White students:** Annual growth rate in program outcomes was higher for African American and Hispanic students in the intervention condition than White students in the intervention condition

Sarama et al., 2012	Randomized Control Trial; Mathematics	Examined interaction between intervention condition and race (African American vs. all other groups)	**Stronger effects for non-White students:** African American students in intervention condition with follow-up demonstrated higher growth in math as compared to African American students in the control condition or intervention condition with no follow up.
Smith, 2001	Non-equivalent control group; Mathematics	Examined interaction between race (White vs. Black) condition	**Equivalent results across groups:** No significant intervention effects for either White or Black students
Stevens & Durkin, 1992	Randomized Control Trial; Literacy	ANCOVA examining interaction between race and treatment	**Equivalent results across groups:** No significant race by treatment interaction
Walton & Cohen, 2011	Randomized Control Trial; Postsecondary Education	Examined interaction between race (European American vs. African American) and condition	**Stronger effects for non-White students:** Intervention improved GPA for African American, but not European American students. Gap in GPA between European and African Americans at beginning of first year was eliminated by end of senior year
Witte et al., 2010	Quasi-Experimental; Comprehensive School Reform	Compared test scores of students in the voucher program to a matched sample of public-school students	**Mixed findings:** Black and Hispanic students demonstrated significantly lower growth in math over 2-year evaluation period than White students. Black students also demonstrated less growth in reading than White students. No significant differences in growth for Asian and Native American students, as compared to White students.
Meta-Analyses			
Cole, 2013	RCT, Quasi-experimental; Literacy	Examined race/ethnicity (Hispanic vs Other and Asian vs. Other) as a moderator of effect sizes	**Equivalent results across groups:** Race/ethnicity was not a significant predictor of effect sizes for oral or written outcomes
Cole, 2014	RCT, Quasi-experimental; Literacy	Examined race/ethnicity (Hispanic vs Other and Asian vs. Other) as a moderator of effect sizes	**Equivalent results across groups:** Race/ethnicity was not a significant predictor of effect sizes
Darrow, 2009	Randomized control trial, quasi-experimental; Literacy	Predicted effect sizes by program ethnicity	**Stronger effects for non-White students:** Effect sizes for programs with non-White racial majority (i.e., predominantly African American, Hispanic, or Asian/Pacific Islander) significantly higher than effect sizes for programs with White majority

Durlak et al., 2011	Study design not indicated; Social-emotional/ behavioral	Compared effect sizes for Caucasian, African American, and Mixed students as rival hypothesis for program effects	**Equivalent results across groups:** Effect sizes not significantly different across groups
Gee, 1995	Experimental; Literacy	ANOVA of program results according to minority status	**Equivalent results across groups:** No significant results
Hill & Tyson, 2009	Interventions and correlational studies; Parental Involvement	Compared strength of relation for European and African American participants	**Equivalent results across groups:** Relationship between involvement and achievement was positive for both African American and European Americans, effects not significantly different
Jeynes, 2007	Study design not indicated; Parental Involvement	Examined differential effect sizes for studies with varying demographics (mostly African American, all African American, mostly Latino & Asian, all Latino & Asian, mostly Asian, all Asian)	**Stronger effects for non-White students:** Effects of parental involvement on academic outcomes were significant for all groups, but largest effects for studies with mostly or all African American participants or mostly or all Asian American participants
Jeynes, 2008	Study design not indicated; Literacy	Compared effect sizes according to % of minority students in the study	**Equivalent results across groups:** Effect sizes were similar for regardless of % minority students in study
Li & Suen, 2012	Experimental or quasi-experimental; Literacy	Predicted effect sizes by ethnicity (Spanish-speaking vs. Other)	**Equivalent results across groups:** Ethnicity did not predict effect sizes
Losinski et al., 2014	Randomized control trial, quasi-experimental; Social-emotional/ behavioral	Predicted effect sizes by percentage African American and percentage White	**Stronger effects for non-White students:** Percentage of African American students significantly, positively influenced treatment effects; Percentage Caucasian had no impact on treatment effects
Manz et al., 2010	Experimental, quasi-experimental, single-subject, qualitative; Literacy	Compared effect sizes between studies with predominantly Caucasian and predominantly minority samples	**Stronger effects for White students:** Effect sizes were moderate-large for studies with predominantly Caucasian samples, but negligible for studies with predominantly minority samples
Springer et al., (1999)	Study design not indicated; Post-secondary education	Compared effect sizes for groups that were predominantly white, predominantly underrepresented minorities, or heterogenous	**Stronger effects for non-White students:** Effects of small-group learning was significant greater for groups composed primarily or exclusively of African American or Latino students compared with predominately white or heterogenous groups

Wilson et al., 2011	RCT, Quasi-experimental, & pre-post; Dropout Prevention	Predicted program odds ratio effect sizes by percent white	**Equivalent results across groups:** Percent white did not predict odds ratio effect sizes