American Psychological Association

Neuropsychology

ISSN: 0894-4105

2021, Vol. 35, No. 2, 141–156
https://doi.org/10.1037/neu0000715

A Systematic Review and Meta-Analysis of Racial/Ethnic Differences and Similarities in Executive Function Performance in the United States

Gianna Rea-Sandin[1], Kevin M. Korous[2], and José M. Causadias[2]
[1] Department of Psychology, Arizona State University
[2] School of Social and Family Dynamics, Arizona State University

OPEN DATA OPEN MATERIALS PREREGISTERED+

Objective: The extent that executive function performance varies between racial/ethnic groups in the United States is unclear, limiting future studies on the problems underlying these differences. The aim of this meta-analysis was to test two competing hypotheses: The cultural differences hypothesis asserts large differences between Whites and racial/ethnic minorities in the U.S., and small differences between- (e.g., African Americans, Latinos) and within- (e.g., Latinos: Mexican Americans, Cuban Americans) minority groups. The cultural similarities hypothesis posits small differences between Whites and minorities, and these differences are equal or smaller in magnitude than differences between- and within-minorities on executive function performance. We also tested moderators of these differences. ***Method:*** We focused on overall executive functioning performance and its three core components: inhibitory control, working memory, and cognitive flexibility.. A systematic search on PsycINFO, Web of Science, ERIC, PubMed, and ProQuest Dissertations and Theses Global identified 46 records (17% unpublished; 38 independent samples) with 56,067 total participants (M_{age} = 44.48 years; range = 3.05–80.45; 52% female; 39.5% racial/ethnic minority). ***Results:*** Absolute differences between Whites and minorities ($d = 0.85$, 95% CI [0.65, 1.05]) were larger in magnitude compared to between-minorities ($d = 0.44$, 95% CI [0.28, 0.60]) and within-minorities ($d = 0.09$, 95% CI [0.03, 0.15]). White-minority differences were moderated by type of executive function measure and year of data collection. Post hoc analyses revealed large relative differences between some groups but not others. ***Conclusions:*** Findings support the cultural differences hypothesis for executive function performance. This meta-analysis underscores the need to address social inequalities in the U.S. that drive performance differences.

Key Points

Question: What is the magnitude of the differences between racial/ethnic groups in the United States on executive functioning performance? ***Findings:*** We found that differences in executive functioning performance were larger between Whites and minorities compared to between- and within-minority groups. ***Importance:*** We argue that these differences are explained by social inequalities experienced by racial/ethnic minorities that compromise their performance rather than inherent differences between groups. ***Next Steps:*** Future research should examine the social factors that account for these differences in performance in an effort to minimize these gaps.

Keywords: meta-analysis, systematic review, executive function, racial/ethnic differences

Supplemental materials: https://doi.org/10.1037/neu0000715.supp

Gianna Rea-Sandin https://orcid.org/0000-0002-3395-7277
Kevin M. Korous https://orcid.org/0000-0001-6825-374X
José M. Causadias https://orcid.org/0000-0001-8249-5497

We acknowledge Sabine Doebel for her feedback on an earlier version of the manuscript. We thank Rylie Fields, Madeline Goldman, Andy Kwon, Mona Said, and Megha Varma for assisting with the systematic search and coding. We also thank Paul Crane, Kimberly Turner Nesbitt, Teresa Seeman, Virginie Patt, Natacha Akshoomoff, William Rodgers, Jeffrey Liew, Aya Williams, Florrie Ng, Whitney Cowell, Carlos Diaz-Venegas, Jim McLean, and Rosemarie Perry for providing the requested information.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We have no conflicts of interest to disclose.

The data are available at https://osf.io/7524g/

The study materials are available at https://osf.io/7524g/

Preregistration Plus Analysis Plan are available at https://psyarxiv.com/tm87j/

Correspondence concerning this article should be addressed to Gianna Rea-Sandin, Department of Psychology, Arizona State University, P.O. Box 871104, Arizona State University, Tempe, AZ 85287, United States. Email: gianna.rea-sandin@asu.edu

Executive functioning is a broad-term indexing a family of effortful top–down mental processes that underlie reasoning, problem-solving, and planning abilities (see Diamond, 2013). Executive functioning is a salient outcome to examine, as it plays a major role across multiple domains, including education and health (Moffitt et al., 2011). In this study, we focus on three components that are well established in the literature: inhibitory control, working memory, and cognitive flexibility. Inhibitory control refers to the ability to deliberately restrain dominant, automatic, or prepotent responses when necessary (Miyake & Friedman, 2012). Working memory is defined as the ability to actively maintain information while performing one or more mental operations, and cognitive flexibility is the ability to shift back and forth between tasks, operations, or mental sets (Diamond, 2013; Miyake & Friedman, 2012).

A growing body of research has examined racial/ethnic differences in executive function performance (e.g., Dore et al., 2015; Flores et al., 2017). However, many studies focus on White-minority comparisons, with less attention to variation between- (e.g., Asian Americans, Latinos) and within- (e.g., Latinos: Mexican Americans, Cuban Americans) minority groups in the United States (Causadias et al., 2018). The goal of this systematic review and meta-analysis was to evaluate the magnitude of differences in executive functioning performance between Whites and minorities and between- and within-minority groups in the United States, as well as the factors that moderate these differences.

Racial/Ethnic Variation in Executive Functioning Performance

In this meta-analysis, the term racial/ethnic minority signifies membership into any non-White cultural groups in the United States. However, we recognize that the term obscures the various ways in which racial/ethnic groups have been minoritized in the U.S. context (e.g., self-exiles vs. involuntary refugees; enslavement vs. colonization; and the disenfranchisement of land or civil rights; Warikoo & Carter, 2009). Nevertheless, this study employed this operationalization of race/ethnicity to synthesize the extant literature. In doing so, this meta-analysis provides more statistical power to detect effects than separate independent studies (Cohn & Becker, 2003) and takes a step toward advancing the field by accounting for research quality, documenting differences in performance across measures, and exposing lack of representation of specific groups across studies.

Research suggests that noncognitive factors related to group membership can affect performance. One study found that Black participants in a stereotype threat condition performed significantly worse on neuropsychological tests compared to Black participants in a nonthreat condition, suggesting that characteristics of the testing environment can obscure the validity of executive functioning measures (Thames et al., 2013). Race-related norms for tests of cognitive performance have been developed to reduce the occurrence of falsely diagnosing cognitive impairment, particularly for Black individuals (Thames et al., 2013). However, there are issues surrounding the use of race-based norms, including the justification for inferior/superior treatment of participants based on racial/ethnic status and failure to address factors that explain differences in performance (Manly, 2005). Together, these issues contribute to the perpetuation of stereotypes toward racial/ethnic minorities.

Despite some notable exceptions (e.g., Thorell et al., 2013), many studies on race/ethnicity and executive functioning center on differences between, and not within, groups (e.g., East–West comparisons; Imada et al., 2013). In many cases, studies that focused on variation in executive functioning in the United States often focus exclusively in Black–White comparisons (e.g., Sisco et al., 2015; Zsembik & Peek, 2001). However, studies that include two or more racial/ethnic minority groups often do not test differences between and within them (e.g., Rodgers et al., 2003).

Differences and Similarities Between and Within-Racial/Ethnic Groups in the United States

The cultural differences hypothesis predicts larger White-minority differences than differences between- (e.g., Blacks, Latinos) and within- (e.g., Latinos: Mexican Americans, Cuban Americans) minority groups on most domains of psychological functioning (Causadias et al., 2018), in this case executive functioning performance. In contrast, the cultural similarities hypothesis predicts White-minority differences of smaller magnitude, while differences between- and within-minority groups are equal or larger in magnitude (Causadias et al., 2018). Some research has found that minorities and Whites varied on executive functioning performance (e.g., Dore et al., 2015; Flores et al., 2017) whereas other studies suggest that executive functioning performance is similar across racial/ethnic groups (e.g., Flores et al., 2017; Merz et al., 2017). Thus, a systematic review and meta-analysis is an optimal method to understand the sources of heterogeneity behind these mixed findings.

To our knowledge, there is no published systematic review or meta-analysis on cultural differences and similarities in executive functioning, although one meta-analysis examining the link between executive functioning and false belief understanding found evidence that children from different countries were more similar than different in terms of executive functioning (Devine & Hughes, 2014). Often, race/ethnicity is treated as a covariate, control, or confound, although there are exceptions (e.g., Rhoades et al., 2011). For instance, Lawson et al. (2018) pooled the results of 25 studies and found that the magnitude of associations between executive functioning and socioeconomic status were small regardless of the racial/ethnic composition of the sample, although diverse samples (r = .17) and predominately White samples (r = .16) were slightly larger than predominately Black samples (r = .06).

Few studies have examined potential moderators of the association between race/ethnicity and executive functioning performance. One study demonstrated that age did not moderate the association between race/ethnicity and rate of change in executive functioning performance (e.g., Wilson et al., 2015). However, the relation between executive functioning performance and racial/ethnic group has been shown to differ by education level, such that higher educational attainment was associated with slower cognitive decline in Black and Latino individuals (e.g., Sheffield & Peek, 2011). In a cross-sectional study, Black participants scored lower on executive functioning than White participants, but these disparities were larger for younger than older participants (Zahodne et al., 2016). Although several studies have demonstrated moderation of the association between race/ethnicity and executive functioning performance, a meta-analytic approach is needed to examine the role of moderators while aggregating across multiple studies.

The Present Study

To address these issues, we conducted a systematic review and meta-analysis of overall executive functioning performance and its three core components: inhibitory control, working memory, and cognitive flexibility. Our goals were to: (a) examine mean-level differences across these components between Whites and minorities, as well as differences between- and within-minority groups in the United States, and (b) to test for potential moderators of these differences (e.g., sex, age, income, education, and measure). We centered on racial/ethnic groups in the United States because it is a unique historical, sociocultural, and educational context for the emergence of differences in executive function performance. This meta-analysis also focuses on task-based assessments. Although more research is needed to validate this claim, it has been suggested that self-reported measures of executive functioning assess personality characteristics rather than cognitive performance (Buchanan, 2016).

We interpreted these findings in terms of support for the cultural differences or similarities hypotheses, following prior meta-analyses of mean-level differences (e.g., Causadias et al., 2018). Primary analyses examined absolute differences between groups, but we also conducted post hoc analyses to examine relative differences. If the absolute average differences between Whites and minorities on mean levels of executive functioning performance were medium (0.36–0.65), large (0.66–1.00), or very large (>1.00), and were larger than the absolute average differences between- and within-minority groups, we considered that these findings support the cultural differences hypothesis. In contrast, if the absolute average differences between Whites and minorities on mean levels of executive functioning performance were very small (0–0.10) or small (0.11–0.35) and were less than or equal to the absolute average differences between- and within-minority groups, we considered that these findings support the cultural similarities hypothesis. We hypothesized that there will be differences between Whites and minorities; however, we anticipated that those differences would be small in magnitude based on previous research (e.g., Merz et al., 2017). We also expected these differences to be equal or smaller than differences between- and within-minority groups, supporting the cultural similarities hypothesis.

This systematic review and meta-analysis was conducted following PRISMA guidelines (Moher et al., 2009). The research protocol was preregistered at Open Science Framework (OSF) (Rea-Sandin et al., 2019) and published in PsyArXiv before coding and data analyses were conducted and completed (https://psyarxiv.com/tm87j/). A set of guidelines for interpreting effect sizes (Funder & Ozer, 2019), different than Cohen's (1992), was published after we preregistered the protocol. We also consider these criteria in the discussion section.

Method

Literature Search

We searched online databases including PsycINFO, Web of Science, ERIC, PubMed, and ProQuest Dissertations and Theses Global to identify an initial number of records. We used the following string: (*cultural* OR *culture* OR *ethnic* OR *ethnicity* OR *race* OR *racial*) AND (*similarity* OR *similarities* OR *sameness* OR *likeness* OR *equivalence* OR *inequality* OR *inequalities* OR *discrepancies* OR *disparity* OR *disparities* OR *dissimilarity* OR *dissimilarities* OR *disproportionately* OR *differentiation* OR *difference* OR *differences* OR *prevalence* OR *incidence* OR *incompatibility*) AND ("*executive function*" OR "*executive dysfunction*" OR *cognitive* OR "*working memory*" OR "*inhibition*" OR "*inhibitory control*" OR "*attention shifting*" OR "*attention focusing*" OR "*sustained attention*" OR "*set shifting*" OR *shifting* OR "*cognitive flexibility*" OR *switching*) AND ("*United States*" OR *U.S.*). When feasible, we restricted the keywords to the title and abstract. Specific search strings and additional restrictions for each database are on OSF. Additionally, we examined (a) the reference lists of eligible records and studies that cited them (i.e., backward and forward search, respectively), (b) titles and abstracts of publications in cognitive and developmental journals over the past 20 years (e.g., *Developmental Psychology, Cognitive Psychology*), and (c) searched for additional unpublished research by examining conference programs over the past 10 years (e.g., *Association for Psychological Science, Society for Research on Child Development*).

Inclusion and Exclusion Criteria

Rayyan Qatar Computing Research Institute (QCRI) (Ouzzani et al., 2016) was used to facilitate the inclusion and exclusion of records, as well as the reliability of included records. The three authors independently evaluated the identified records using the criteria below. Discrepancies were resolved as a team to reach a complete agreement on the records included.

Variables

We included records that reported sufficient information to compute differences for two or more racial/ethnic groups on at least one component of executive functioning (i.e., inhibitory control, working memory, cognitive flexibility, or a composite of executive functioning, as formed by the primary study). We included records with measures of executive functioning that were performance-based (e.g., go/no-go, Stroop task). We excluded records with self-reported measures of executive functioning, studies that focused on planning and problem-solving, and records that examined self-regulation, cognitive ability/performance, executive attention, memory, emotion regulation, or cognitive decline because these outcomes are related to, but distinct from, executive functioning because introducing additional components, on top of the typical examples of executive functioning, would increase the heterogeneity in our overall mean-level differences. Including higher-order functions would also limit our ability to explain the variability attributed to specific components and measures because higher-order functions often utilize multiple components of executive functioning (Karr et al., 2018).

Participants and Geographic Location

We included records with normative samples regardless of participants' age, sex, culture, race, and ethnicity. However, we excluded records that focused only on one racial/ethnic group and samples collected outside of the United States, records that compared U.S. and non-U.S. samples, records focused on clinical samples (e.g., patients with Alzheimer's disease, schizophrenia patients) or special populations (e.g., gifted children and pregnant adolescents), and records that compared normative and clinical samples. However, we included records drawn from community samples to examine the correlation of other symptoms, such as sleep disturbances or alcohol use, with executive functioning. We excluded records that generated data using computer models (e.g., Morrison et al., 2011).

Language

We only included records written in English.

Research Design

We included records regardless of publication status. We only included cross-sectional and longitudinal studies.

Time Period

We did not impose eligibility criteria based on the date of publication.

Animal Studies

We only included records that used human participants.

Coding

Reliability

We created a coding manual that captured effect sizes and relevant study characteristics. The three authors and five trained research assistants coded 16 records that met initial inclusion for full-text review. Discrepancies were resolved and the coding descriptions were refined accordingly. The remaining records after full-text review were coded independently by two of the authors. We calculated Cohen's Kappa (κ) and Intraclass Correlations (ICC) as measures of interrater reliability between the two coders for categorical and continuous data, respectively (Orwin & Vevea, 2009). After reliability was computed, the two coders resolved discrepancies.

Mean Group Differences

In addition to sample size, we coded for means and standard deviations for executive functioning, disaggregated by race/ethnicity. If this information was not provided, we coded the reported standardized mean difference (e.g., Cohen's *d*, Hedges' *g*) either between two racial/ethnic groups (e.g., Whites compared to Latinos) or between Whites and minorities. If needed, we coded odds ratios, *t*-tests, or *F*-tests and χ^2 tests with one degree of freedom. Regression coefficients, risk ratios, and hazard ratios were not coded. There was good reliability between the two coders on the effect size metric (κ = 1.0) and their respective values (ICC = .93–1.0). We contacted authors if there was insufficient data to compare at least two racial/ethnic groups on at least one component of executive functioning, and records were excluded if the authors we contacted did not provide these data.

Moderators

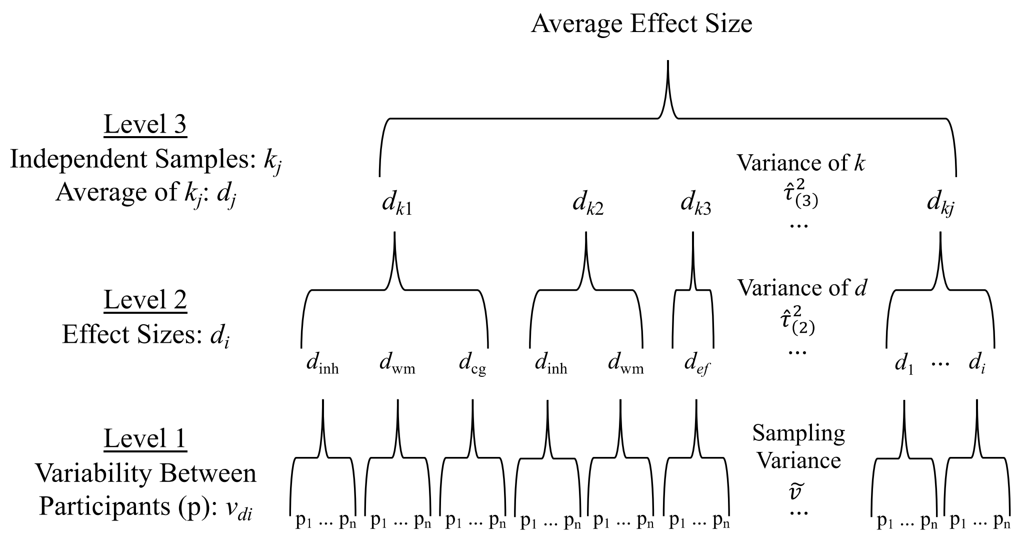
First, we coded the mean age of the sample for a particular comparison (e.g., between Whites and minorities; ICC = .89). Second, we coded the sex composition (i.e., % female) of the sample for a particular comparison (ICC = .90). Third, we coded the mean annual income (ICC = .92) and years of education (ICC = .90) as components of socioeconomic status. For the latter, we coded the number of years and contacted authors when studies reported education categorically (e.g., the level of education). We coded whether socioeconomic status was low, high, or mixed, however, reliability was poor (κ = .18). Only 4% of the records directly described their sample composition, thus socioeconomic composition was dropped as a variable from analyses. Fourth, we coded the component of executive functioning performance including inhibition, working memory, cognitive flexibility, or overall executive functioning (e.g., multiple components; κ = .78). Fifth, we coded the type of performance-based measure of executive functioning was coded (e.g., Go/No-Go task; κ = .73). Sixth, we coded the year of data collection (ICC = .99). Lastly, we coded whether the record was published or unpublished (i.e., dissertation and conference presentation; κ = 1.0).

Study Treatment and Analyses

We computed Cohen's *d* and its variance following formulas by Borenstein et al. (2009). We computed effect sizes as positive to indicate the magnitude of the *absolute* difference between groups. Thus, effect sizes indicate the average difference in mean levels of executive functioning performance. To estimate the overall difference, we employed a random-effects model. In effect, studies with larger sample sizes were given more weight because they often provide more precise estimates (Borenstein et al., 2009). As described later, we conducted post hoc analyses to examine *relative* differences between groups (i.e., the average direction of the difference). We conducted a structural equation modeling based meta-analysis (for a review, see Cheung, 2015a). This approach is similar to conducting a regression analysis in a structural equation framework, but the goal is to predict an effect size rather than an outcome. Meta-analyses are traditionally a two-level model, but we specified a third level to account for dependencies introduced by extracting multiple effect sizes from a single record (Cheung, 2014). For instance, some records could have included effect sizes for multiple components of executive functioning performance, for different measures of performance, or for different subgroups of their sample (see Figure 1). Likewise, multiple records could have used data from the same sample of participants (e.g., national data set), but vary in their inclusion of participants. In each of these cases, we extracted each of the effect sizes to increase the amount of information about the distribution of the population of effect sizes. Nonetheless, because these effect sizes overlap in the participants included, they are clustered within independent samples. We used *k* to denote independent samples, which is our level-three unit of analysis (or cluster variable). Individual effect sizes extracted from each *k* were included as our level-two unit of analysis and the variability of each effect size, which is assumed to be known

Figure 1

Three-Level, Random-Effects Model. In This Example, Three Effect Sizes (i.e., One for Each Component of Executive Functioning) Were Extracted From Sample One (d_{k1})*, Two From Sample Two* (d_{k2})*, One From Sample Three* (d_{k3})*. The Number of Effect Sizes Extracted From Each Independent Sample Can Vary.* v_{di} *= Assumed Known Variance of Each Extracted Effect Size; inh = Inhibition; wm = Working Memory; cg = Cognitive Flexibility; ef = Composite Executive Functioning, n = the nth Participant; i = the ith Effect Size; and j = the jth Independent Sample. This Figure by the Authors is Licensed Under a Creative Commons Attribution 4.0 International License*



and used to weight effect sizes, is considered as level one in our three-level meta-SEM.

All analyses were conducted in the R platform (version 3.5.0) using RStudio (version 1.1.453) with the metaSEM package (version 1.1.0; Cheung, 2015b). We estimated the average difference (i.e., intercept) for each comparison and its 95% confidence interval (95% CI). We also reported the systematic heterogeneity (I^2) and estimated population variance ($\hat{\tau}^2$) for each level. Because I^2 and $\hat{\tau}^2$ were estimated, we provided likelihood-based CIs (LBCI; Casella & Berger, 2002; Cheung, 2014). We compared models in which we restricted $\hat{\tau}^2$ at each level to zero and constrained $\hat{\tau}^2$ at each level to be equal. Following recommendations (Cheung, 2015a), we also estimated our model using Maximum-Likelihood (ML) estimation and Restricted Maximum-Likelihood (REML) estimation. As a sensitivity test for outliers, we excluded effect sizes that were larger than 2.5 absolute deviations below or above the median (Leys et al., 2013), calculated in the psych package (version 1.8.12; Revelle, 2018). For this, we estimated the median effect size, which was used to calculate the median absolute deviation. We report the median effect size and median absolute deviation for these analyses in the supplemental materials.

We included moderators as covariates in our three-level model, which attempts to explain the variability between effect sizes at level two ($\hat{\tau}^2_{(2)}$) and between-samples at level three ($\hat{\tau}^2_{(3)}$). Moderators were treated as predictors similar to covariates in a conventional structural equation model. We tested each moderator in separate models. Moderators were tested in separate models and were treated as predictors. Continuous moderators (i.e., mean age, sex composition, mean education, and year of data collection) were grand-mean centered (i.e., average value across the independent samples) to ease interpretation. The intercept reflected the average difference for studies around the grand mean whereas the slope indicated the difference in the average effect size for each unit increase above the grand mean. To test the moderating role of the component of executive functioning and performance-based measure of executive functioning, we computed a dummy code for each category. For example, all effect sizes associated with inhibitory control were coded as one and all other components of executive functioning were coded as zero. We included all dummy codes for a particular moderator in the model and constrained the intercept to zero, which estimated the average effect size for each dummy code (Cheung, 2015a). We tested whether the effect size varied between categories by testing the difference between the model with and without the dummy codes. All effect sizes informed our moderator analyses even if they had missing data because we conducted analyses with and without Full Information Maximum-Likelihood Estimation (FIML). We tested the assumption of missing at random by comparing the overall magnitude between effect sizes with complete and missing moderator values. There were no significant differences.

Publication Bias

First, we tested the publication status of records as a moderator. The slope in this analysis indicated the difference in the average effect between published and unpublished records. Second, we ran Egger's regression (Egger et al., 1997), which is a more common test of publication bias (Jin et al., 2015), utilizing the metabias command of the meta-package in the R platform (version 4.9-2; Schwarzer, 2007). Third, we used the selection method within the weightr package (version 1.1.2; Coburn & Vevea, 2017) in the R platform. In this approach, we specified a priori weights representing moderate and severe one-tailed selection (Vevea & Woods, 2005). Importantly, we used a three-level model for our test of moderation, which accounted for the dependency between effect sizes. The last two tests, on the other hand, were developed under the assumption of independence.

Results

Literature Search

Our search process was conducted in August 2018. Our search string identified 5,836 unique records from online databases and excluded 5,730 records for reasons listed in Figure 2. The remaining 106 records were considered for full-text review. Additional methods identified 71 extra records, totaling 177 full-texts for review.

We contacted 58 corresponding authors via email and nine of those authors responded with relevant statistics.[1] We also contacted authors of eight conference abstracts in which three provided relevant statistics as published records or unpublished data. Overall, 46 records fully met our criteria for inclusion; 38 published, seven dissertations, and one unpublished conference data. From these 46 records, we identified 38 independent samples (k) to include in our meta-analysis. The number of independent samples is smaller than the number of records because some records reported on an overlapping sample of participants. We excluded records as duplicates when they reported the same relevant statistics from the same sample of participants (e.g., when means and standard deviations were the same between the two records). Additionally, two records contributed more than one independent sample; one record reported means separately for males and females and another record reported information from four separate studies. For our three-level analyses reported below, k is level three, effect sizes extracted from each k is level two, and the variance of each effect size is level one. Furthermore, we refer to between-study heterogeneity as the variability between effect sizes at level three (k) and within-study heterogeneity as the variability between effect sizes at level two, or within k.

Study Characteristics

Across the samples ($k = 38$), we computed 290 effects sizes (range per k: 1–48) differences in executive functioning performance between Whites and minorities ($k = 35$, 206 effect sizes), between-minorities ($k = 11$, 84 effect sizes), and within-minorities ($k = 2$, 7 effect sizes). Some records provided information for all three comparisons, thus, the k across the comparisons is larger than the total k. The distributions of effect sizes are compared in a stem-and-leaf plot on OSF, as well as forest plots.

The sample size across studies ranged from 36 to 18,892 participants (mean sample size = 831) with approximately 56,067 participants in total, which includes White (60.5%), Black (23.8%), Latino (12.9%), Asian American (1.6%), multiracial (0.2%), and Pacific Islander (0.02%) participants within the United States. For 1% of non-White participants, we could not extract a specific breakdown. The average percentage of females across studies was 52% (range = 0%–100%) and the average mean age was 44.48 years (range = 3.05–80.45), however we could not extract the mean age for every sample. Finally, the year of data collection ranged from 1996 to 2012, although we only identified the year of data collection for 29% of the studies.

Absolute Differences Between Whites and Minorities

Overall Absolute Difference

We extracted 206 effect sizes ($k = 35$) that represent the difference between White and minority participants on executive functioning performance. Of these, 47.6% were between White and Black participants, 36.4% were between White and Latino participants, 5.8% were between White and Asian American participants, 1.5% were between White and Pacific Islander participants, and 8.7% were between White and non-White (i.e., more than one racial/ethnic group) participants.

Our estimated absolute difference between Whites and minorities was $d = 0.85$, 95% CI [0.65, 1.05], $\hat{\tau}^2_{(2)} = .400$, 95% LBCI [.317, .509], $\hat{\tau}^2_{(3)} = .226$, 95% LBCI [.110, .449]. We found that 63.15% of the total variability was attributed to within-study heterogeneity ($I^2_{(2)} = 63.15$, 95% LBCI [45.48, 78.71]), whereas 35.72% was attributed to between-study heterogeneity ($I^2_{(3)} = 35.72$, 95% LBCI [20.02, 53.64]). We could not reject the null hypothesis of equal variance in our multilevel model [$\chi^2(1) = 2.30$, $p = .129$]. That is, the variability at both levels was not significantly different (i.e., $\hat{\tau}^2_{(2)} = \hat{\tau}^2_{(3)}$). Additionally, the estimated population variability within [$\chi^2(1) = 5562.92$, $p < .001$] and between [$\chi^2(1) = 38.19$, $p < .001$] studies were significantly different from zero. Our results based on ML estimation were not biased compared to REML estimation, $d = 0.85$, 95% CI [0.65, 1.05], $\hat{\tau}^2_{(2)} = .400$, $\hat{\tau}^2_{(3)} = .236$, $I^2_{(2)} = 62.21$, $I^2_{(3)} = 36.67$.

As a sensitivity test, we removed 15 effect sizes larger than 2.5 deviations, leaving 191 effects sizes ($k = 35$). The absolute effect size was attenuated and the population variability was also reduced, $d = 0.70$, 95% CI [0.57, 0.82], $\hat{\tau}^2_{(2)} = .161$, 95% LBCI [.122, .213], $\hat{\tau}^2_{(3)} = .079$, 95% LBCI [.031, .173]. However, the amount of systematic variability explained by level-two and level-three remained comparable, $I^2_{(2)} = 65.20$, 95% LBCI [45.17, 82.64], $I^2_{(3)} = 31.95$, 95% LBCI [14.27, 40.10].

Moderation of Absolute Difference

We report the results of our moderator analyses in Table 1. We did not find evidence of moderation by mean age ($k = 33$, 198 effect sizes), sex composition ($k = 35$, 206 effect sizes), or mean education ($k = 18$, 124 effect sizes) of racial/ethnic differences in executive function performance. When we used FIML to include missing values, we found similar results for mean age (est. = 0.004, 95% CI [−0.002, 0.01]) and mean education (est. = −0.05, 95% CI [−0.16, 0.07]). We also found that the magnitude of absolute differences was similar across components [$\chi^2(3) = 4.99$, $p = .172$].

[1] One record utilized data from the Alzheimer's Disease Neuroimaging Initiative (ADNI; http://adni.loni.usc.edu/) and was unable to directly share the data with us, therefore we derived means/*SDs* from ADNI directly.

Figure 2
PRISMA Flow Diagram of Search Procedure

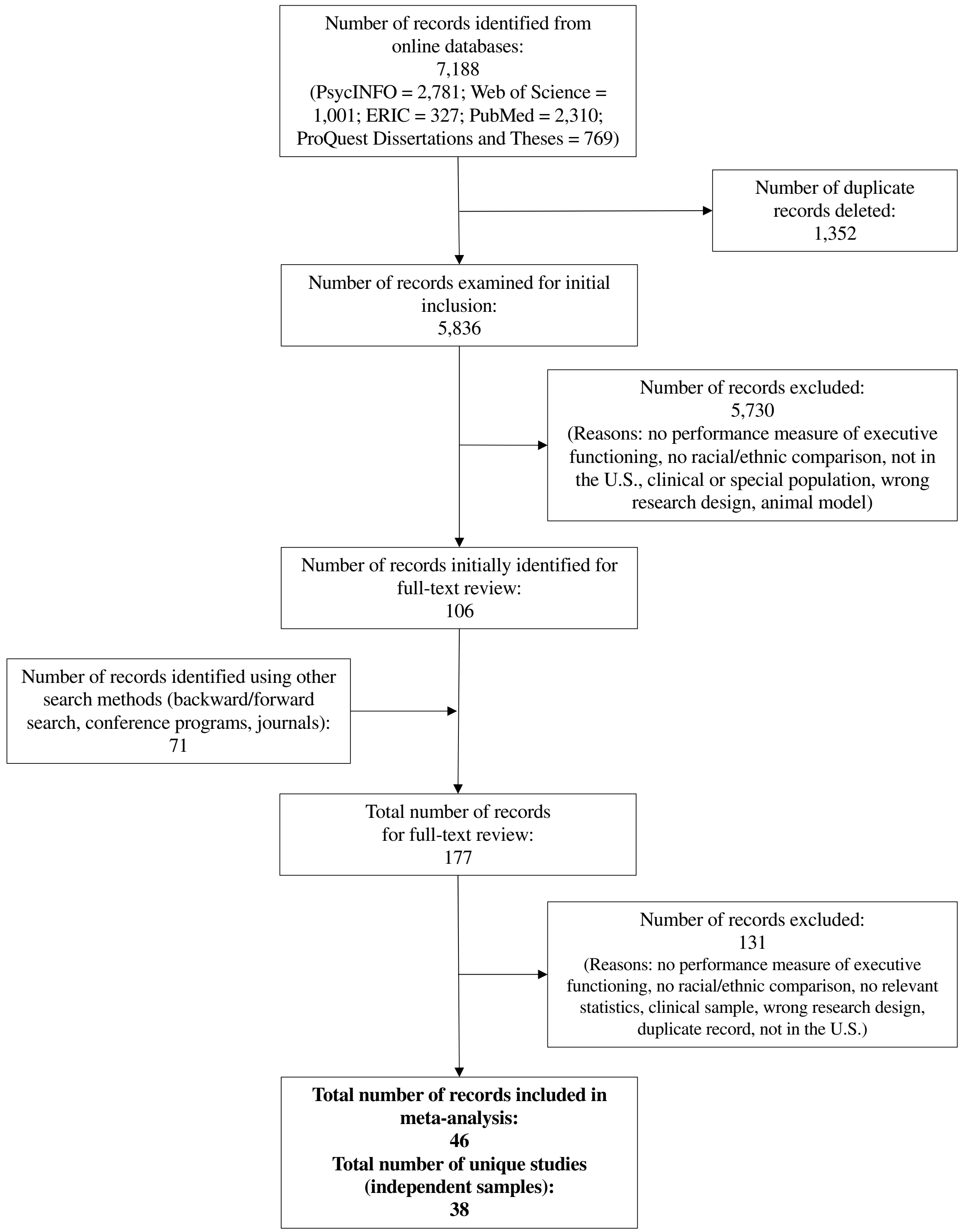


Table 1
Summary of Moderators for Absolute Differences Between Whites and Minorities (Cohen's d) on Executive Function Performance

	k	Est.	p value	[95% CI]	$\hat{\tau}^2_{(2)}$	$\hat{\tau}^2_{(3)}$	$R^2_{(2)}$	$R^2_{(3)}$
Mean age	33				.416	.217	.001	.071
Intercept at 45.71 years		0.87	.000	[0.66, 1.07]				
Slope		0.01	.193	[−0.002, 0.01]				
Sex composition (% female)	35				.400	.219	.000	.029
Intercept at 51% female		0.85	.000	[0.65, 1.05]				
Slope		0.51	.428	[−0.76, 1.79]				
Mean education	18				.456	.358	.002	.027
Intercept at 13.22 years		0.99	.000	[0.67, 1.31]				
Slope		−0.06	.445	[−0.23, 0.10]				
EF (executive functioning) component*					.398	.189	.004	.166
Inhibition	8	0.74	.000	[0.37, 1.12]				
Working memory	15	0.65	.000	[0.35, 0.94]				
Cognitive flexibility	20	1.01	.000	[0.78, 1.25]				
EF composite	9	0.85	.000	[0.54, 1.16]				
Performance-based measure*					.005	.007	.433	.000
Continuous performance	2	0.21	.662	[−0.74, 1.17]				
Digit span	10	0.65	.002	[0.25, 1.05]				
Wisconsin card sorting	2	1.35	.000	[0.91, 1.79]				
Fluency tasks	13	0.95	.000	[0.63, 1.26]				
Stroop tasks	3	0.82	.010	[0.20, 1.45]				
Color trails	10	1.23	.000	[0.88, 1.60]				
Flanker tasks	2	0.21	.505	[−0.41, 0.82]				
Dimensional card sorting	2	0.17	.598	[−0.45, 0.78]				
Composite scores	15	0.91	.000	[0.62, 1.20]				
Year of data collection	11				.350	.035	.046	.576
Intercept at 2006		0.83	.000	[0.61, 1.06]				
Slope		−0.06	.018	[−0.10, −0.01]				

Note. k = number of independent samples; Est. = estimate of average association (d) or slope (Beta); 95% CI = 95% confidence interval; $\hat{\tau}^2_{(2)}$ and $\hat{\tau}^2_{(3)}$ = variability within- and between-study, respectively. $R^2_{(2)}$ and $R^2_{(3)}$ = proportion of within- and between-variability explained by including predictor compared to no predictor. Intercept indicates the average effect size (d) at the specified value whereas the slope (Beta) represents the change in the average effect size for each increase above the intercept.
* = Intercept free model.

To test moderation by performance-based measures, we selected measures that were included in at least two independent samples. Of the 184 effect sizes that met this inclusion, 71 represented composite scores of executive functioning composed of more than one performance-based measure, as computed by the authors of each study. The remaining 113 effect sizes represented fluency tasks (35%), color trails (17%), Wisconsin card sorting (17%), digit span (15%), flanker task (5%), dimensional card sorting (5%), Stroop tasks (4%), and continuous performance tasks (CPT; 2%). We found that the magnitude of absolute differences was significantly different between the performance-based measures [$\chi^2(8) = 20.18$, $p = .010$]. We explained 11.1% of level-two variability and 2.5% of level-three variability by allowing the effect sizes to vary between performance-based measures.

We also found evidence of moderation by the year of data collection ($k = 11$, 40 effect sizes). That is, for every increase in the year of data collection the magnitude of absolute differences in EF performance decreased by 0.06. The year of data collection explained 57.6% of level-three variability and 4.6% of level-two variability. Given the amount of missing values for the year of data collection we estimated a model using FIML and found similar results (intercept = 0.88, 95% CI [0.66, 1.10], est. = −0.07, 95% CI [−0.12, −0.03], $p = .002$, $R^2_{(2)} = .027$, $R^2_{(3)} = .479$).

Overall, we found that performance-based measures and the year of data collection significantly accounted for some of the variations in the magnitude of absolute differences. For this reason, we included both moderators in one model to test if they remained significant when controlling for the other. However, our model would not converge, which is likely related to the amount of missing values for the year of data collection ($k = 11$, 40 effect sizes).

Publication Bias

Of the 206 effect sizes ($k = 35$), we extracted 21 effect sizes from unpublished sources ($k = 8$). We did not find a significant difference in the magnitude of effect size between published and unpublished records [$\chi^2(1) = 3.07$, $p = .080$, est. = −0.43, $p = .075$]. Our Egger's test was also not significant, $t(204) = 1.18$, $p = .240$. However, when we specified cutoff points and a priori weights, the overall effect was attenuated. We specified four intervals of p values to ensure adequate coverage: Less than .005 (130 effect sizes), between .005 and .050 (22 effect sizes), between .050 and .250 (33 effect sizes), and greater than .250 (21 effect sizes). Under moderate one tail-selection weights the effect size was $d = 0.71$ and under severe one tail-selection weights the effect size was $d = 0.56$, compared to the baseline model ($d = 0.84$).

Absolute Differences Between-Minorities

Overall Absolute Difference

We extracted 84 effect sizes ($k = 11$) that represent the difference between-minority participants on EF performance. Of these, 64.3%

were between Black and Latino participants, 11.9% were between Latino and Asian American participants, 8.3% were between Black and Asian American participants, 3.6% were between Black and Pacific Islander participants, and 3.6% were between Latino and Pacific Islander participants.

Our estimated absolute difference between-minorities was $d = 0.44$, 95% CI [0.28, 0.60], $\hat{\tau}^2_{(2)} = .207$, 95% LBCI [.134, .320], $\hat{\tau}^2_{(3)} = .033$, 95% LBCI [.002, .154]. We found that 82.30% of the total variability was attributed to within-study heterogeneity ($I^2_{(2)} = 82.30$, 95% LBCI [53.48, 100.00]), whereas 13.20% was attributed to between-study heterogeneity ($I^2_{(3)} = 13.20$, 95% LBCI [0.00, 43.14]). We rejected the null hypothesis of equal variance [$\chi^2(1) = 5.01$, $p = .025$]. That is, the variability was significantly different between the two levels. Additionally, the estimated population variability within [$\chi^2(1) = 245.77$, $p < .001$] and between [$\chi^2(1) = 3.16$, $p = .076$] studies were significantly different from zero. Our results based on ML estimation were not biased compared to REML estimation, $d = 0.44$, 95% CI [0.27, 0.61], $\hat{\tau}^2_{(2)} = .207$, $\hat{\tau}^2_{(3)} = .041$, $I^2_{(2)} = 79.97$, $I^2_{(3)} = 15.66$. Likewise, in this model we rejected the null hypothesis of equal variance between level two and level three [$\chi^2(1) = 3.95$, $p = .047$].

As a sensitivity test, we removed 10 effect sizes larger than 2.5 deviations, leaving 74 effects sizes ($k = 11$). The absolute effect size was attenuated and the population variability was also reduced, $d = 0.30$, 95% CI [0.21, 0.39], $\hat{\tau}^2_{(2)} = .014$, 95% LBCI [.004, .030], $\hat{\tau}^2_{(3)} = .016$, 95% LBCI [.003, .027]. In addition, the amount of systematic variability explained by level-two was substantially reduced, $I^2_{(2)} = 33.13$, 95% LBCI [10.03, 63.96], whereas level-three variability increased, $I^2_{(3)} = 38.39$, 95% LBCI [7.65, 72.10].

Moderation of Absolute Difference

We report the results of our moderator analyses in Table 2. We did not find evidence of moderation by mean age ($k = 11$, 84 effect sizes), sex composition ($k = 11$, 84 effect sizes), or mean education ($k = 4$, 34 effect sizes) of racial/ethnic differences in executive function performance. When we used FIML to include missing values, we found similar results for mean education (est. = −0.10, 95% CI [−0.25, 0.05]). We also found that the magnitude of absolute differences was similar across components [$\chi^2(3) = 0.99$, $p = .804$].

To test moderation by performance-based measures, we selected measures that were included in at least two independent samples. However, very few of the 11 independent samples used 2 of the same tasks. Therefore, we grouped effect sizes into two categories; those that were based on individual performance-based measures compared to those based on multiple measures (composite scores). We specified composite scores as the reference category so that the slope represented the change in magnitude for effect sizes based on individual measures, although it was not significant. The slope for the year of data collection was also not significant ($k = 4$, 16 effect sizes). Our model using FIML would not converge, which is likely related to the amount of missing values for the year of data collection (81%).

Publication Bias

We only identified one effect size from an unpublished study. Therefore, we could not test the publication status as a moderator. Our Egger's test was not significant, $t(82) = 0.55$, $p = .584$. However, when we specified cutoff points and a priori weights the overall effect was attenuated. We specified four intervals of p values to

Table 2

Summary of Moderators for Absolute Differences Between-Minorities (Cohen's d) on Executive Function Performance

	k	Est.	p value	[95% CI]	$\hat{\tau}^2_{(2)}$	$\hat{\tau}^2_{(3)}$	$R^2_{(2)}$	$R^2_{(3)}$
Mean age	11				.208	.032	.000	.029
Intercept at 35.20 years		0.44	.000	[0.27, 0.61]				
Slope		−0.000	.938	[−0.01, 0.01]				
Sex Composition (% female)	11				.202	.034	.024	.000
Intercept at 49% female		0.46	.000	[0.30, 0.63]				
Slope		0.94	.407	[−1.29, 3.18]				
Mean education	4				.101	.053	.018	.401
Intercept at 13.90 years		0.52	.000	[0.24, 0.81]				
Slope		−0.16	.128	[−0.37, 0.05]				
EF component*					.206	.028	.005	.160
Inhibition	23	0.53	.000	[0.28, 0.78]				
Working memory	20	0.44	.002	[0.15, 0.72]				
Cognitive flexibility	16	0.46	.003	[0.15, 0.76]				
EF composite	25	0.34	.015	[0.07, 0.62]				
Performance-based measure					.205	.028	.007	.152
Intercept at composite scores	5	0.34	.005	[0.10, 0.58]				
Slope at individual tasks	8	0.15	.289	[−0.13, 0.44]				
Year of data collection	4				.029	.079	.014	.188
Intercept at 2008		0.58	.000	[0.27, 0.88]				
Slope		−0.02	.359	[−0.07, 0.03]				

Note. k = number of independent samples; Est. = estimate of average association (d) or slope (Beta); 95% CI = 95% confidence interval; $\hat{\tau}^2_{(2)}$ and $\hat{\tau}^2_{(3)}$ = variability within- and between-study, respectively. $R^2_{(2)}$ and $R^2_{(3)}$ = proportion of within- and between-variability explained by including predictor compared to no predictor. Intercept indicates the average effect size (d) at the specified value whereas the slope (Beta) represents the change in the average effect size for each increase above the intercept.
* = Intercept free model.

ensure adequate coverage: Less than .005 (33 effect sizes), between .005 and .050 (12 effect sizes), between .050 and .250 (21 effect sizes), and greater than .250 (18 effect sizes). Under moderate one tail-selection weights the effect size was $d = 0.35$ and under severe one tail-selection weights the effect size was $d = 0.24$, compared to the baseline model ($d = 0.45$).

Absolute Differences Within-Minorities

Overall Absolute Difference

We extracted seven effect sizes ($k = 2$) that represent the differences within-minority participants on EF performance. One study contributed six effect sizes between Mexican, Puerto-Rican, Cuban, and Other Latino participants in the United States (Díaz-Venegas et al., 2019) and another study contributed one effect size between Mexican and Dominican participants in the United States (Ng et al., 2015). We did not identify any studies looking at within-minority differences in EF performance among Blacks, Asian Americans, Pacific Islanders, or Native Americans.

Because there were only two independent samples and seven effect sizes, we could not estimate a three-level random-effects model nor could we run moderation analyses or tests of publication bias. Alternatively, we estimated the overall absolute difference within-minorities using a two-level fixed-effects model. Our estimated absolute difference within-minorities was $d = 0.09$, 95% CI [0.03, 0.15]. The test of heterogeneity based on the Q statistic was nonsignificant [$\chi^2(6) = 1.98, p = .921$]. Importantly, this effect size is specific to the studies included and the confidence interval may be wider than we estimated (Schmidt et al., 2009).

Post Hoc Analyses—Relative Differences

We conducted post hoc tests for overall relative differences between groups to better understand the magnitude of the differences we found. We did not provide specific hypotheses given the exploratory nature of these analyses. In contrast with absolute differences in which numerical signs are ignored, relative differences consider positive (+) and negative (−) signs to make comparisons. These allowed us to determine which groups performed higher than others when the direction of performance is considered. We provide a detailed description of these analyses and results in the Supplemental material, as well as on OSF.

In sum, we found that White participants had higher scores on executive functioning performance than minority participants, $d = 0.78$ (excluding outliers, $d = 0.62$). We did not find evidence of moderation by mean age, sex composition, mean education, or components of executive functioning on racial/ethnic differences in executive function performance. However, we did find evidence of moderation by the year of data collection. We also found some evidence of publication bias (see Supplemental material). Table 3 includes the results for the meta-analysis of each racial/ethnic comparison.

Discussion

To our knowledge, this is the first systematic review and meta-analysis to evaluate the extent to which racial/ethnic groups in the United States differ on executive functioning performance, as well as moderators of these differences. Overall, based on conventions by Cohen (1992) and Sawilowsky (2009), we found large absolute differences between Whites and minorities ($d = .85$), medium-sized differences between-minority groups ($d = .44$), and very small differences within-minority groups ($d = .09$; Figure 3). We found that the magnitude of absolute differences between Whites and minorities was moderated by the type of performance-based measure and year of data collection. Together, these findings support the cultural differences hypothesis that asserts large differences in executive functioning performance between Whites and minorities in the United States and smaller differences in performance between- and within-minority groups. Post hoc analyses revealed a consistent pattern of differences between groups. Figure 4 illustrates all relative differences in order of magnitude. In terms of executive functioning performance, Whites and Asian Americans

Table 3

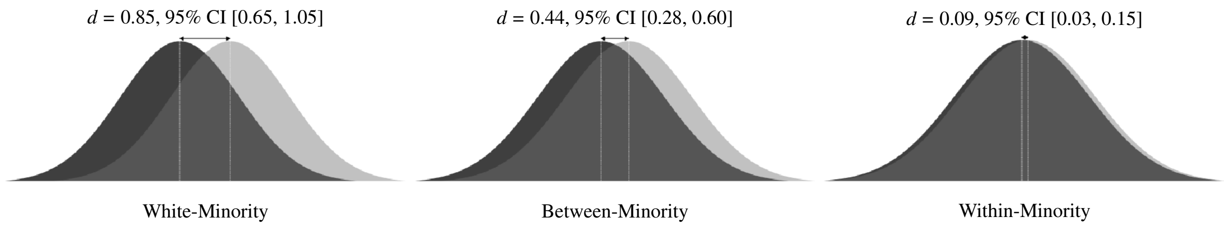
Summary of Absolute and Relative Differences Between Black, Latino, Asian American, and White Participants (Cohen's d) on Executive Function Performance

			Absolute differences							Relative differences						
Comparison	k	No. of ES	d	p value	[95% CI]	$\hat{\tau}^2_{(2)}$	$\hat{\tau}^2_{(3)}$	$I^2_{(2)}$	$I^2_{(3)}$	d	p value	[95% CI]	$\hat{\tau}^2_{(2)}$	$\hat{\tau}^2_{(3)}$	$I^2_{(2)}$	$I^2_{(3)}$
Black–Latino	11	54	0.33	.000	[0.19, 0.47]	.158	.008	89.44	4.64	0.03	.778	[−0.20, 0.26]	.242	.054	78.91	17.67
White–Asian	6	12	0.33	.000	[0.14, 0.51]	.071	.000	79.34	0.00	0.18	.188	[−0.09, 0.45]	.129	.025	74.73	14.63
Asian–Latino	4	10	0.71	.000	[0.33, 1.10]	.350	.000	94.83	0.00	0.62	.006	[0.17, 1.06]	.478	—	96.16	—
Asian–Black	4	7	0.69	.000	[0.37, 1.01]	.000	.088	0.00	83.64	0.69	.000	[0.37, 1.01]	.000	.088	0.00	83.64
White–Black	25	98	0.86	.000	[0.60, 1.12]	.322	.289	52.25	46.79	0.80	.000	[0.52, 1.08]	.417	.332	55.24	43.98
White–Latino	12	75	0.93	.000	[0.63, 1.23]	.570	.107	82.57	15.54	0.84	.000	[0.49, 1.18]	.746	.124	84.47	14.05

Note. k = number of independent samples; No. of ES = number of effect sizes; d = average standardized mean difference (Cohen's d) in executive function performance; 95% CI = 95% confidence interval; $\hat{\tau}^2_{(2)}$ and $\hat{\tau}^2_{(3)}$ = variability within- and between-study, respectively. $I^2_{(2)}$ and $I^2_{(3)}$ = the percentage of systematic variability between effect sizes explained by level-two or level-three, respectively. For relative differences positive effect sizes (d) indicate that the sample to the left of the dash mark (under the comparison column) had higher executive function performance levels than the sample to the right of the dash mark. Rows are organized in terms of the magnitude of the relative difference. The double dash indicates that we constrained level-three variability to zero in order to estimate our model.

Figure 3

Graphic Depiction of our Findings on White-Minority, Between-Minority, and Within-Minority Differences in Executive Function Performance. Findings Support the Cultural Differences Hypothesis



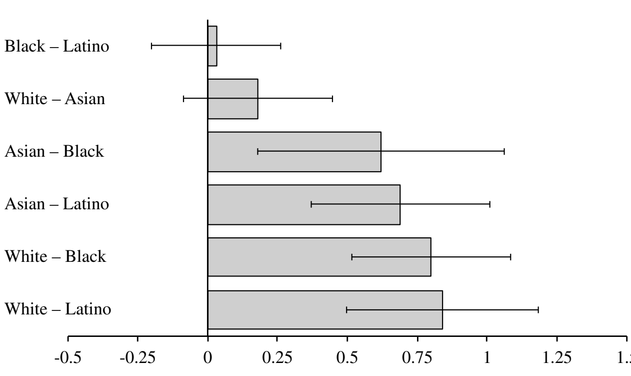
were more similar than different and had higher scores than Latinos and Blacks, who were also more similar than different in this domain of functioning.

There are several factors that could explain racial/ethnic differences in executive functioning performance. Stereotype threat can have negative repercussions in executive functioning performance. Several studies including Black participants demonstrated that higher levels of reported racial discrimination were associated with poorer performance on neuropsychological tests, and this is due to the allocation of cognitive resources to regions involved in emotion regulation, thus resulting in poorer performance on these tasks (e.g., Masten et al., 2011; Thames et al., 2013).

In the United States, racial/ethnic minorities must cope with racism in addition to other stressors (Levy et al., 2016). Clark et al. (1999) argued that health disparities and differences in psychological functioning can be understood by examining the effect of racism in stress and coping. Race-based social stress can affect the physiological activity and sleep, influencing educational outcomes through its effects on important cognitive processes like executive functioning (Levy et al., 2016). In one study, higher levels of cortisol were related to lower executive functioning performance, particularly for Black individuals compared to White (Blair et al., 2011). Cortisol is an indicator of stress neurobiology, a potential mechanism by which adverse social experiences due to social inequality perpetuates differences in cognitive functioning (Blair et al., 2011), although its link with racial discrimination is more complex than previously expected (Korous et al., 2017). In families, racial disparities in health and psychological functioning may be amplified by the intergenerational effects of social inequality on physiological stress response systems (Lu & Halfon, 2003).

Figure 4

Overall Relative Differences (Cohen's d) for Specific Racial/Ethnic Comparisons in Order of the Magnitude of the Effect Size. Positive Effect Sizes Indicate That the Sample to the Left of the Dash Mark Had Higher Executive Function Performance Levels Than the Sample to the Right of the Dash Mark. The Horizontal Line for Each Bar Represents Their 95% CI



In addition, research has suggested that other factors, such as linguistic ability and acculturation influence executive functioning task performance, potentially explaining why we found large differences between Whites and racial/ethnic minorities (see O'Bryant et al., 2004). One study examining the influence of sociodemographic variables on neuropsychological test performance in Spanish-speaking older adults found that acculturation was the best predictor of digit span performance (Acevedo et al., 2007), highlighting the importance that language and cultural practices have on executive functioning task performance. However, race/ethnicity, native language, and acculturation are rarely reported in research studies, limiting our ability to examine them as potential confounds or covariates (O'Bryant et al., 2004).

We also found important variability within- and between-studies. That is, effect sizes extracted from the same sample of participants varied in their magnitude and/or direction (within-study), which could be attributed to methodological factors such as different components or measures of executive functioning performance. Additionally, effect sizes between-samples also varied in their magnitude and/or direction (between-study), which could be attributed to sample characteristics such as age or year of data collection. Interestingly, for the absolute difference between Whites and minorities on executive function performance, we found that there was just as much variability within-samples as there was between-samples. Almost all of this variability (over 90%) was systematic, meaning that sample and method decisions have some influence on the magnitude and/or direction of absolute differences between Whites and minorities.

Among the moderators we tested, we found that the type of executive function performance task and the year of data collection explained some of this variability, which we discuss in more detail below. Nonetheless, future meta-analyses on executive functioning performance should consider the different levels of variability so it is not attributed solely to between-study variance. When we examined specific absolute group comparisons in our post hoc analyses, we found similar amounts of within-sample variability for most comparisons. If we attribute within-sample variability to the type of racial/ethnic comparison, then there appears to be more variability between effect sizes associated with differences between-minorities than effect sizes associated with differences between Whites and minorities. One possible explanation for this finding is that there is more variability within- and between-minority groups than what is often acknowledged (Causadias et al., 2018; Coll et al., 1996). The underlying assumption behind grouping together individuals with a common racial or ethnic background is that they share similar social experiences (Okazaki & Sue, 1995). Nevertheless, combining racial/ethnic minorities into a single group can be misleading and potentially uninformative because it neglects within-group heterogeneity; thus, alternative systems of racial classification should be considered (see Bonilla-Silva, 2004).

Racial/ethnic differences in executive functioning performance did not vary by age. Although all age groups were represented, a majority of the primary studies comprised older adult populations. To better capture the development of executive functioning across the lifespan, future primary studies should target samples of children and adolescents from different racial/ethnic groups in the United States. Importantly, these findings do not suggest that there is no variation in executive function performance across ages. Instead, these findings suggest that *differences* in executive functioning performance between racial/ethnic groups are observed across all age groups. The fact that these gaps in performance do not vary significantly across age says little about their emergence and continuity. Future research should employ longitudinal designs to elucidate the origin and trajectories of these differences in performance, and the correlates and mechanisms behind them. For instance, to what degree is there overlap with trajectories observed for socioeconomic differences in executive functioning performance (Hackman et al., 2015).

Another important finding was that absolute differences in performance were consistent across inhibitory control, working memory, and cognitive flexibility components, concordant with previous research. For example, a systematic review and meta-analysis examining differences in executive functioning performance for patients with major depressive disorder and healthy control patients found similar effect sizes across domains of executive functioning (Snyder, 2013). In contrast, Lawson et al. (2018) found that the effect sizes on the correlation between socioeconomic status and executive functioning performance varied across the domains in their meta-analysis. However, they tested moderation using an analysis of variance approach that did not account for the dependencies between effect sizes, whereas we employed meta-regression that accounted for the nonindependence.

We found that effect sizes between Whites and minorities differed by type of performance-based measure. Absolute differences were larger than zero for the Wisconsin card sorting task, color trails tasks, fluency tasks, composite measures of executive functioning, Stroop tasks, and digit span, whereas the 95% CI for differences between CPT, flanker tasks, and the dimensional card sorting task included zero. The tasks most sensitive to group differences comprised the majority of the measures represented in this study compared to tasks that did not show group differences, potentially explaining why findings differed by task but not by a component of executive functioning. It is also possible that differences could be attributed to measurement properties of a task, rather than actual differences in ability. Measurement characteristics of a task are assumed to be invariant across participants (Meredith, 1993) yet the majority of normed executive functioning tasks are based on White monolingual Americans (Gasquoine, 2009). This can be a problem because it could lead to the perpetuation of deficiency models of racial/ethnic minorities as well as the misdiagnosis of minorities as having cognitive deficits when they are neurologically intact (Gasquoine, 2009).

Although one potential solution is to develop race-based norms for executive functioning tasks, this comes with other issues, including the inferior treatment of racial/ethnic minorities, overlooking other factors that could contribute to performance (e.g., acculturation, racism, and socioeconomic status), and the heterogeneity that exists among racial/ethnic groups (Gasquoine, 2009). One alternative solution is the individual comparison standard approach (Gasquoine, 2009), which estimates neurological skill levels on a case-by-case basis by considering demographic variables (that are differentially weighted), vocabulary test scores, reading level, and the highest score obtained on the battery of executive functioning tasks (Lezak et al., 2004). In addition to improving the interpretability of executive functioning task performance, this approach also has the potential to improve clinicians' ability to assess an individual's cognitive strengths and limitations (Mitrushina et al., 2005). However, for clinicians who are taxed with a high volume of work, administering a battery of executive functioning performance tasks may not be feasible. In those cases, we encourage clinicians to employ measures that are less likely to demonstrate racial/ethnic differences, such as The Dimensional Change Card Sort (DCCS; Zelazo, 2006) or CPT (e.g., Test of Variables of Attention; Leark et al., 2007).

We found that as the year of data collection increased, the magnitude of absolute differences in executive functioning performance decreased between Whites and minorities. In our post hoc analyses, we did not find a link between the year of data collection and the magnitude of the relative differences. However, when we estimated a model using FIML given the amount of missing data, the year of data collection was a significant moderator. Together, these findings suggest that White-minority differences in executive functioning performance are narrowing over time. These findings are consistent with evidence showing that compared to earlier cohorts, later cohorts performed better on tests of cognitive functioning (e.g., Christensen et al., 2013). In fact, Sheffield and Peek (2011) found that the narrowing racial/ethnic differences in cognitive impairment were due to larger absolute decreases in the prevalence of cognitive impairment for Blacks and Latinos than for Whites.

Importantly, we found some evidence of publication bias, potentially explaining why we found absolute and relative differences between White and minority participants of a magnitude that can be interpreted as large (Cohen, 1992; Sawilowsky, 2009) or grossly overestimated (Funder & Ozer, 2019). There is reason to believe that not all effect sizes of the differences between groups were included across these studies (e.g., subgroup reporting bias; Hahn et al., 2000). For example, 71% of the 84 effect sizes between-minorities were nonsignificant whereas only 24% of the 246 effect sizes between Whites and minorities were nonsignificant. It is possible that published studies comparing Whites and minorities are likely to report significant group differences whereas within-minority studies are likely to report nonsignificant group differences. Although both comparisons are likely impacted by this bias, this still provides evidence for the cultural differences hypothesis.

This study has several limitations. First, we were unable to test some of our study aims since we identified a small number of effect sizes, particularly within-minority groups. Studies included in this meta-analysis were 61% White, 24% Black, 13% Latino, and 2% Asian American, multiracial, and Pacific Islander. There were no comparisons with Native American participants in any of the included studies. This dramatic underrepresentation of Asian Americans, Native Americans, and Pacific Islanders calls into question the extent to which these findings capture differences between- and within-racial/ethnic groups in the United States. This is consistent with criticisms of the invisibility of minority participants and researchers in American psychology (Syed et al., 2018). For example, Filipino Americans are the third-largest group of Asian Americans, yet they are grossly underrepresented in psychological research (Nadal, 2011). Future research should include more minority participants, not just for the sake of inclusion, but to be able to make authoritative claims about human behavior.

Additionally, these findings should be replicated in other large countries with multiethnic populations. For instance, investigators in China can examine the magnitude of differences in executive function performance between the dominant group (Han Chinese) and minorities (e.g., Uyghur, Tibetan, and Zhuang), as well as between- and within-minorities. In order for future research in other countries to overcome the limitations of the current study, researchers should report more detailed sample characteristics, including income, education, occupation, and overall socioeconomic status, and assess executive function performance with measures created, translated, normed, and/or validated for their specific populations.

Second, we could not test the role of several moderators due to the lack of available data. For instance, we could not estimate a multiple regression model including the type of performance-based measures and year of data collection, likely due to the amount of missing data for the year of data collection. Therefore, we could not rule out if they are confounded. Future primary studies should include thorough sample characteristics to more accurately understand the moderators of the association between race/ethnicity and executive functioning performance. We could not fully explain the amount of systematic heterogeneity, therefore, there could be unexamined moderators.

Third, we were not able to document the multiple cultural processes that drive these differences, restricting our understanding of executive functioning performance across racial/ethnic groups. We did examine years of education as one component of socioeconomic status, but we failed to find moderation for absolute or relative differences between racial/ethnic groups. We were not able to test levels of socioeconomic status more broadly as a moderator because these data were not consistently reported in most studies included in the meta-analysis. In a recent meta-analysis on socioeconomic status and executive functioning performance, racial composition (i.e., >60% Black, White, or Mixed samples) was not a significant moderator (Lawson et al., 2018). Importantly, future studies should consider examining both socioeconomic status and race/ethnicity as predictors in the same model. Our analysis, as well as Lawson et al. (2018), considers each as a moderator instead of examining the contribution of each variable over the other. A meta-analysis on this subject could employ meta-analytic structural equation models to estimate path models that tackle this question (Cheung, 2015a). We expect both to contribute to variability in executive function performance, because racial/ethnic differences persist across levels of socioeconomic status (e.g., Assari et al., 2019).

Despite these challenges, this is the first systematic review and meta-analysis to test the cultural differences and similarities hypothesis in executive functioning performance across the lifespan. An important strength of this study was the examination of both absolute and relative differences in executive functioning performance, a more objective approach to help minimize the use of deficit models when interpreting differences between groups while also acknowledging racial inequalities. Finally, we calculated the majority of effect sizes using means and standard deviations, as opposed to relying on reported effect sizes. In doing so, we obtained more consistent estimates. We also employed a multilevel meta-SEM approach, which maximized the amount of included evidence that we identified in this literature.

In sum, our study contributes to the literature by illuminating the magnitude of differences between and within-racial/ethnic groups. Executive functioning skills underlie adaptive functioning across multiple domains. Therefore, it is imperative for future research to examine the cultural factors that explain these differences and to move away from conceptualizations that explain them in terms of the inferiority of racial/ethnic minority groups (Okazaki & Sue, 1995). Here, rather than arguing that there is something inherently dysfunctional about the executive functioning performance of some racial/ethnic groups compared to others, we argue that the problem lies within the social inequalities that drive these differences.

References

Acevedo, A., Loewenstein, D. A., Agrón, J., & Duara, R. (2007). Influence of sociodemographic variables on neuropsychological test performance in Spanish-speaking older adults. *Journal of Clinical and Experimental Neuropsychology*, *29*, 530–544. https://doi.org/10.1080/13803390600814740

*Akshoomoff, N., Newman, E., Thompson, W. K., McCabe, C., Bloss, C. S., Chang, L., Amaral, D. G., Casey, B. J., Ernst, T. M., Frazier, J. A., Gruen, J. R., Kaufmann, W. E., Kenet, T., Kennedy, D. N., Libiger, O., Mostofsky, S., Murray, S. S., Sowell, E. R., Schork, N., & Jernigan, T. L. (2014). The NIH Toolbox Cognition Battery: Results from a large normative developmental sample (PING). *Neuropsychology*, *28*, 1–10. https://doi.org/10.1037/neu0000001

Assari, S., Caldwell, C. H., & Bazargan, M. (2019). Association between parental educational attainment and youth outcomes and role of race/ethnicity. *JAMA Network Open*, *2*, e1916018–e1916018. https://doi.org/10.1001/jamanetworkopen.2019.16018

*Blair, C., Granger, D. A., Willoughby, M., Mills-Koonce, R., Cox, M., Greenberg, M. T., Kivlighan, K. T., Fortunato, C. K., & the FLP Investigators. (2011). Salivary cortisol mediates effects of poverty and parenting on executive functions in early childhood. *Child Development*, *82*, 1970–1984. https://doi.org/10.1111/j.1467-8624.2011.01643.x

Bonilla-Silva, E. (2004). From bi-racial to tri-racial: Towards a new system of racial stratification in the USA. *Ethnic and Racial Studies*, *27*, 931–950. https://doi.org/10.1080/0141987042000268530

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

*Bosco, F., Allen, D. G., & Singh, K. (2015). Executive attention: An alternative perspective on general mental ability, performance, and subgroup differences. *Personnel Psychology*, *68*, 859–898. https://doi.org/10.1111/peps.12099

*Brewster, P. W., Melrose, R. J., Marquine, M. J., Johnson, J. K., Napoles, A., MacKay-Brandt, A., Farias, S., Reed, B., & Mungas, D. (2014). Life experience and demographic influences on cognitive function in older adults. *Neuropsychology*, *28*, 846–858. https://doi.org/10.1037/neu0000098

Buchanan, T. (2016). Self-report measures of executive function problems correlate with personality, not performance-based executive function measures, in nonclinical samples. *Psychological Assessment*, *28*, 372–385. https://doi.org/10.1037/pas0000192

*Buckhalt, J. A., El-Sheikh, M., & Keller, P. (2007). Children's sleep and cognitive functioning: Race and socioeconomic status as moderators of effects. *Child Development*, *78*, 213–231. https://doi.org/10.1111/j.1467-8624.2007.00993.x

*Burke, S. L., Rodriguez, M. J., Barker, W., Greig-Custo, M. T., Rosselli, M., Loewenstein, D. A., & Duara, R. (2018). Relationship between cognitive performance and measures of neurodegeneration among Hispanic and White non-Hispanic individuals with normal cognition, mild cognitive impairment, and dementia. *Journal of the International Neuropsychological Society*, *24*, 176–187. https://doi.org/10.1017/S1355617717000820

Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Brooks/Cole.

*Caughy, M. O. B., Mills, B., Owen, M. T., & Hurst, J. R. (2013). Emergent self-regulation skills among very young ethnic minority children: A confirmatory factor model. *Journal of Experimental Child Psychology*, *116*, 839–855. https://doi.org/10.1016/j.jecp.2013.07.017

Causadias, J. M., Korous, K. M., & Cahill, K. M. (2018). Are Whites and minorities more similar than different? Testing the cultural similarities hypothesis on psychopathology with a second order meta-analysis. *Development and Psychopathology*, *30*, 2009–2027. https://doi.org/10.1017/S0954579418000895

Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, *19*, 211–229. https://doi.org/10.1037/a0032968

Cheung, M. W.-L. (2015a). *Meta-analysis: A structural equation modeling approach*. Wiley.

Cheung, M. W.-L. (2015b). MetaSEM: An R package for meta-analysis using structural equation modeling. *Frontiers in Psychology*, *5*, 1–7. https://doi.org/10.3389/fpsyg.2014.01521

Christensen, K., Thinggaard, M., Oksuzyan, A., Steenstrup, T., Andersen-Ranberg, K., Jeune, B., McGue, M., & Vaupel, J. W. (2013). Physical and cognitive functioning of people older than 90 years: A comparison of two Danish cohorts born 10 years apart. *The Lancet*, *382*, 1507–1513. https://doi.org/10.1016/S0140-6736(13)60777-1

Clark, R., Anderson, N. B., Clark, V. R., & Williams, D. R. (1999). Racism as a stressor for African Americans. A biopsychosocial model. *American Psychologist*, *54*, 805–816. https://doi.org/10.1037/0003-066X.54.10.

Coburn, K. M., & Vevea, J. L. (2017). *weigthr: Estimating weight-function models for publication bias*. https://CRAN.R-project.org/package=weightr

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cohn, L. D., & Becker, B. J. (2003). How meta-analysis increases statistical power. *Psychological Methods*, *8*, 243–253. https://doi.org/10.1037/1082-989X.8.3.243

Coll, C. G., Crnic, K., Lamberty, G., Wasik, B. H., Jenkins, R., Garcia, H. V., & McAdoo, H. P. (1996). An integrative model for the study of developmental competencies in minority children. *Child Development*, *67*, 1891–1914. https://doi.org/10.2307/1131600

*Cowell, W., Rauh, V., Margolis, A., & Herbstman, J. (2017, April). *Effects of persistent and transient material hardship on executive function: Evidence from a low-income, minority birth cohort* [Poster presentation]. Society for Research in Child Development, Austin, Texas.

*Crane, P. K., Narasimhalu, K., Gibbons, L. E., Pedraza, O., Mehta, K. M., Tang, Y., Manly, J. J., Reed, B. R., & Mungas, D. M. (2008). Composite scores for executive function items: Demographic heterogeneity and relationships with quantitative magnetic resonance imaging. *Journal of the International Neuropsychological Society*, *14*, 746–759. https://doi.org/10.1017/S1355617708081162

Devine, R. T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Development*, *85*, 1777–1794. https://doi.org/10.1111/cdev.12237

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

*Díaz-Venegas, C., Downer, B., Langa, K. M., & Wong, R. (2019). Cognitive functioning of US adults by race and Hispanic origin. In W. A. Vega, J. L. Angel, L. M. F. G. Robledo, & K. S. Markides, (Eds.), *Contextualizing health and aging in the Americas* (pp. 85–107). Springer. https://doi.org/10.1007/978-3-030-00584-9_5

Dore, G. A., Waldstein, S. R., Evans, M. K., & Zonderman, A. B. (2015). Associations between diabetes and cognitive function in socioeconomically diverse African Americans and whites. *Psychosomatic Medicine*, *77*, 643–652. https://doi.org/10.1097/PSY.0000000000000196

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

*Fazeli, P. L., Ross, L. A., Vance, D. E., & Ball, K. (2013). The relationship between computer experience and computerized cognitive test performance among older adults. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, *68*, 337–346. https://doi.org/10.1093/geronb/gbs071

*Fitzpatrick, A. L., Rapp, S. R., Luchsinger, J., Hill-Briggs, F., Alonso, A., Gottesman, R., Lee, H., Carnethon, M., Liu, K., Williams, K., Sharrett, A. R., Frazier-Wood, A., Lyketsos, C., & Seeman, T. (2015). Sociodemographic correlates of cognition in the Multi-Ethnic Study of Atherosclerosis (MESA). *The American Journal of Geriatric Psychiatry*, *23*, 684–697. https://doi.org/10.1016/j.jagp.2015.01.003

Flores, I., Casaletto, K. B., Marquine, M. J., Umlauf, A., Moore, D. J., Mungas, D., Gershon, R. C., Beaumont, J. L., & Heaton, R. K. (2017). Performance of Hispanics and Non-Hispanic Whites on the NIH Toolbox Cognition Battery: The roles of ethnicity and language backgrounds. *The Clinical Neuropsychologist*, *31*, 783–797. https://doi.org/10.1080/13854046.2016.1276216

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*, 156–168. https://doi.org/10.1177/2515245919847202

Gasquoine, P. G. (2009). Race-norming of neuropsychological tests. *Neuropsychology Review*, *19*, 250–262. https://doi.org/10.1007/s11065-009-9090-5

*Gladsjo, J. A., Schuman, C. C., Evans, J. D., Peavy, G. M., Miller, S. W., & Heaton, R. K. (1999). Norms for letter and category fluency: Demographic corrections for age, education, and ethnicity. *Assessment*, *6*, 147–178. https://doi.org/10.1177/107319119900600204

Hahn, S., Williamson, P. R., Hutton, J. L., Garner, P., & Flynn, E. V. (2000). Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. *Statistics in Medicine*, *19*, 3325–3336. https://doi.org/10.1002/1097-0258(20001230)19:24<3325::AID-SIM827>3.0.CO;2-D

Hackman, D. A., Gallop, R., Evans, G. W., & Farah, M. J. (2015). Socioeconomic status and executive function: Developmental trajectories and mediation. *Developmental Science*, *18*, 686–702. https://doi.org/10.1111/desc.12246

*Held, M. M. (2013). *The Effects of gender, culture, and acculturation on verbal fluency in Filipino Americans and Caucasian European Americans* [Doctoral dissertation]. Palo Alto University.

Imada, T., Carlson, S. M., & Itakura, S. (2013). East–West cultural differences in context-sensitivity are evident in early childhood. *Developmental Science*, *16*, 198–208. https://doi.org/10.1111/desc.12016

Jin, Z. C., Zhou, X. H., & He, J. (2015). Statistical methods for dealing with publication bias in meta-analysis. *Statistics in Medicine*, *34*, 343–360. https://doi.org/10.1002/sim.6342

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin*, *144*, 1147–1185. https://doi.org/10.1037/bul0000160

*Kaufman, A. S., McLean, J. E., & Reynolds, C. R. (1991). Analysis of WAIS-R factor patterns by sex and race. *Journal of Clinical Psychology*, *47*, 548–557.
*Kemmotsu, N. (2010). *Performance of Japanese Americans on selected cognitive instruments* [Doctoral dissertation]. UC San Diego.
Korous, K. M., Causadias, J. M., & Casper, D. M. (2017). Racial discrimination and cortisol output: A meta-analysis. *Social Science & Medicine*, *193*, 90–100. https://doi.org/10.1016/j.socscimed.2017.09.042
*LaRue, A., Romero, L. J., Ortiz, I. E., Chi Lang, H., & Lindeman, R. D. (1999). Neuropsychological performance of Hispanic and non-Hispanic older adults: An epidemiologic survey. *Clinical Neuropsychologist*, *13*, 474–486. https://doi.org/10.1076/1385-4046(199911)13:04;1-Y;FT474
Lawson, G. M., Hook, C. J., & Farah, M. J. (2018). A meta-analysis of the relationship between socioeconomic status and executive function performance among children. *Developmental Science*, *21*, 1–22. https://doi.org/10.1111/desc.12529
Leark, R. A., Greenberg, L. M., Kindschi, C. L., Dupuy, T. R., & Hughes, S. J. (2007). *Test of variables of attention continuous performance test*. The TOVA Company.
Levy, D. J., Heissel, J. A., Richeson, J. A., & Adam, E. K. (2016). Psychological and biological responses to race-based social stress as pathways to disparities in educational outcomes. *American Psychologist*, *71*, 455–473. https://doi.org/10.1037/a0040322
Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*, 764–766. https://doi.org/10.1016/j.jesp.2013.03.013
Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment* (4th ed.). Oxford University Press.
*Liew, J., Chen, Q., & Hughes, J. N. (2010). Child effortful control, teacher–student relationships, and achievement in academically at-risk children: Additive and interactive effects. *Early Childhood Research Quarterly*, *25*, 51–64. https://doi.org/10.1016/j.ecresq.2009.07.005
*Louie, J. Y. (2013). *The influence of culture and adoption status on the development of behavioral inhibition, anxiety, temperament, and parental control* [Doctoral dissertation]. UCLA.
Lu, M. C., & Halfon, N. (2003). Racial and ethnic disparities in birth outcomes: A life-course perspective. *Maternal and Child Health Journal*, *7*, 13–30. https://doi.org/10.1023/A:1022537516969
*Lundy, S. M. (2007). *The relationship between anxious/depressed and withdrawn symptoms on cognitive and academic measures in elementary school children*. ProQuest Dissertations Publishing.
Manly, J. J. (2005). Advantages and disadvantages of separate norms for African Americans. *The Clinical Neuropsychologist*, *19*, 270–275. https://doi.org/10.1080/13854040590945346
*Manly, J. J., Jacobs, D. M., Sano, M., Bell, K., Merchant, C. A., Small, S. A., & Stern, Y. (1998). Cognitive test performance among nondemented elderly African Americans and whites. *Neurology*, *50*, 1238–1245. https://doi.org/10.1212/WNL.50.5.1238
*Manly, J. J., Jacobs, D. M., Touradji, P., Small, S. A., & Stern, Y. (2002). Reading level attenuates differences in neuropsychological test performance between African American and White elders. *Journal of the International Neuropsychological Society*, *8*, 341–348. https://doi.org/10.1017/S1355617702813157
Masten, C. L., Telzer, E. H., & Eisenberger, N. I. (2011). An fMRI investigation of attributing negative social treatment to racial discrimination. *Journal of Cognitive Neuroscience*, *23*, 1042–1051. https://doi.org/10.1162/jocn.2010.21520
*Menon, C., Hall, J., Hobson, V., Johnson, L., & O'Bryant, S. E. (2012). Normative performance on the executive clock drawing task in a multiethnic bilingual cohort: A project FRONTIER study. *International Journal of Geriatric Psychiatry*, *27*, 959–966. https://doi.org/10.1002/gps.2810
Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543. https://doi.org/10.1007/BF02294825
*Merz, E. C., Landry, S. H., Montroy, J. J., & Williams, J. M. (2017). Bidirectional associations between parental responsiveness and executive function during early childhood. *Social Development*, *26*, 591–609. https://doi.org/10.1111/sode.12204
Mitrushina, M., Boone, K. B., Razani, J., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment* (2nd ed.). Oxford University Press.
Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*, 8–14. https://doi.org/10.1177/0963721411429458
Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thompson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 2693–2698. https://doi.org/10.1073/pnas.1010076108
Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, *151*, 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135
Morrison, R. G., Doumas, L. A., & Richland, L. E. (2011). A computational account of children's analogical reasoning: Balancing inhibitory control in working memory and relational representation. *Developmental Science*, *14*, 516–529. https://doi.org/10.1111/j.1467-7687.2010.00999.x
*Mungas, D., Reed, B. R., Farias, S. T., & DeCarli, C. (2009). Age and education effects on relationships of cognitive test scores with brain structure in demographically diverse older persons. *Psychology and Aging*, *24*, 116–128. https://doi.org/10.1037/a0013421
Nadal, K. L. (2011). *Filipino American psychology: A handbook of theory, research, and clinical practice*. Wiley. https://doi.org/10.1002/9781118094747
*Napier, D. (2014). *Predicting adolescents' academic achievement: The contribution of attention and working memory*. ProQuest Dissertations Publishing.
*Nesbitt, K. T., Baker-Ward, L., & Willoughby, M. T. (2013). Executive function mediates socio-economic and racial differences in early academic achievement. *Early Childhood Research Quarterly*, *28*, 774–783. https://doi.org/10.1016/j.ecresq.2013.07.005
*Ng, F. F. Y., Tamis-LeMonda, C., Yoshikawa, H., & Sze, I. N. L. (2015). Inhibitory control in preschool predicts early math skills in first grade: Evidence from an ethnically diverse sample. *International Journal of Behavioral Development*, *39*, 139–149. https://doi.org/10.1177/0165025414538558
O'Bryant, S., O'Jile, J., & McCaffrey, R. (2004). Reporting of demographic variables in neuropsychological research: Trends in the current literature. *The Clinical Neuropsychologist*, *18*, 229–233. https://doi.org/10.1080/13854040490501439
Okazaki, S., & Sue, S. (1995). Methodological issues in assessment research with ethnic minorities. *Psychological Assessment*, *7*, 367–375. https://doi.org/10.1037/1040-3590.7.3.367
Orwin, R. G., & Vevea, J. L. (2009). Evaluating coding decisions. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 178–200). Russell Sage Foundation.
Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—A web and mobile app for systematic reviews. *Systematic Reviews*, *5*, 210–219. https://doi.org/10.1186/s13643-016-0384-4
*Patt, V. M., Brown, G. G., Thomas, M. L., Roesch, S. C., Taylor, M. J., & Heaton, R. K. (2018). Factor analysis of an expanded Halstead-Reitan Battery and the structure of neurocognition. *Archives of Clinical Neuropsychology*, *33*, 79–101. https://doi.org/10.1093/arclin/acx043
*Perry, R. E., Braren, S. H., Blair, C., & The Family Life Project Key Investigators. (2018). Socioeconomic risk and school readiness: Longitudinal mediation through children's social competence and executive

function. *Frontiers in Psychology*, *9*, Article 1544. https://doi.org/10.3389/fpsyg.2018.01544

*Philbrook, L. E., Hinnant, J. B., Elmore-Staton, L., Buckhalt, J. A., & El-Sheikh, M. (2017). Sleep and cognitive functioning in childhood: Ethnicity, socioeconomic status, and sex as moderators. *Developmental Psychology*, *53*, 1276–1285. https://doi.org/10.1037/dev0000319

*Proctor, A., & Zhang, J. (2008). Performance of three racial/ethnic groups on two tests of executive function: Clinical implications for traumatic brain injury (TBI). *NeuroRehabilitation*, *23*, 529–536. https://doi.org/10.3233/NRE-2008-23609

*Razani, J., Burciaga, J., Madore, M., & Wong, J. (2007). Effects of acculturation on tests of attention and information processing in an ethnically diverse group. *Archives of Clinical Neuropsychology*, *22*, 333–341. https://doi.org/10.1016/j.acn.2007.01.008

Rea-Sandin, G., Korous, K. M., & Causadias, J. M. (2019, October 31). *A systematic review and meta-analysis of cultural differences and similarities in executive functioning in the United States*. https://osf.io/7524g/?view_only=74e2af58188c4a359c53ab165b88bbb4

Revelle, M. (2018). *Psych: Procedures for psychological, psychometric, and personality research*. https://personality-project.org/r/psych-manual.pdf

*Rhoades, B. L., Greenberg, M. T., Lanza, S. T., & Blair, C. (2011). Demographic and familial predictors of early executive function development: Contribution of a person-centered perspective. *Journal of Experimental Child Psychology*, *108*, 638–662. https://doi.org/10.1016/j.jecp.2010.08.004

Rodgers, W. L., Ofstedal, M. B., & Herzog, A. R. (2003). Trends in scores on tests of cognitive ability in the elderly US population, 1993–2000. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *58*, S338–S346. https://doi.org/10.1093/geronb/58.6.S338

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods; JMASM*, *8*, 597–599. https://doi.org/10.22237/jmasm/1257035100

Schmidt, F. L., Oh, I.-S., & Hayes, T. L. (2009). Fixed-versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology*, *62*, 97–128. https://doi.org/10.1348/000711007X255327

*Schneider, B. C., Gross, A. L., Bangen, K. J., Skinner, J. C., Benitez, A., Glymour, M. M., Sachs, B. C., Shih, R. A., Sisco, S., Manly, J. J., & Luchsinger, J. A. (2015). Association of vascular risk factors with cognition in a multiethnic sample. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, *70*, 532–544. https://doi.org/10.1093/geronb/gbu040

Schwarzer, G. (2007). Meta: An R package for meta-analysis. *R News*, *7*, 40–45. https://cran.r-project.org/doc/Rnews/Rnews_2007-3.pdf

*Seeman, T. E., Miller-Martinez, D. M., Stein Merkin, S., Lachman, M. E., Tun, P. A., & Karlamangla, A. S. (2011). Histories of social engagement and adult cognition: Midlife in the US study. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, *66*(Suppl. 1), 141–152. https://doi.org/10.1093/geronb/gbq091

*Shaked, D. (2016). *The relation of socioeconomic status and executive function: Is prefrontal cortex volume a mediator?* University of Maryland.

Sheffield, K. M., & Peek, M. K. (2011). Changes in the prevalence of cognitive impairment among older Americans, 1993–2004: Overall trends and differences by race/ethnicity. *American Journal of Epidemiology*, *174*, 274–283. https://doi.org/10.1093/aje/kwr074

*Sisco, S., Gross, A. L., Shih, R. A., Sachs, B. C., Glymour, M. M., Bangen, K. J., Benitez, A., Skinner, J., Schneider, B. C., & Manly, J. J. (2015). The role of early-life educational quality and literacy in explaining racial disparities in cognition in late life. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, *70*, 557–567. https://doi.org/10.1093/geronb/gbt133

Snyder, H. R. (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: A meta-analysis and review. *Psychological Bulletin*, *139*, 81–132. https://doi.org/10.1037/a0028727

*Staton, L., Hinnant, J. B., Buckhalt, J., & El-Sheikh, M. (2014). Sleep and cognitive performance: The role of income and respiratory sinus arrhythmia reactivity. *Developmental Psychobiology*, *56*, 1528–1540. https://doi.org/10.1002/dev.21247

Syed, M., Santos, C., Yoo, H. C., & Juang, L. P. (2018). Invisibility of racial/ethnic minorities in developmental science: Implications for research and institutional practices. *American Psychologist*, *73*, 812–826. https://doi.org/10.1037/amp0000294

Thames, A. D., Hinkin, C. H., Byrd, D. A., Bilder, R. M., Duff, K. J., Mindt, M. R., Arentoft, A., & Streiff, V. (2013). Effects of stereotype threat, perceived discrimination, and examiner race on neuropsychological performance: Simple as black and white? *Journal of the International Neuropsychological Society*, *19*, 583–593. https://doi.org/10.1017/S1355617713000076

Thorell, L. B., Veleiro, A., Siu, A. F., & Mohammadi, H. (2013). Examining the relation between ratings of executive function and academic achievement: Findings from a cross-cultural study. *Child Neuropsychology*, *19*, 630–638. https://doi.org/10.1080/09297049.2012.727792

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods*, *10*, 428–443. https://doi.org/10.1037/1082-989X.10.4.428

Warikoo, N., & Carter, P. (2009). Cultural explanations for racial and ethnic stratification in academic achievement: A call for a new and improved theory. *Review of Educational Research*, *79*, 366–394. https://doi.org/10.3102/0034654308326162

*Weissberger, G. H., Salmon, D. P., Bondi, M. W., & Gollan, T. H. (2013). Which neuropsychological tests predict progression to Alzheimer's disease in Hispanics? *Neuropsychology*, *27*, 343–355. https://doi.org/10.1037/a0032399

*Williams, A. I., Uchikoshi, Y., Bunge, S. A., & Zhou, Q. (2019). Relations of English and Heritage Language Proficiency to Response Inhibition and Attention Shifting in Dual Language Learners in Head Start. *Early Education and Development*, *30*, 357–374. https://doi.org/10.1080/10409289.2018.1544812

*Wilson, R. S., Capuano, A. W., Sytsma, J., Bennett, D. A., & Barnes, L. L. (2015). Cognitive aging in older Black and White persons. *Psychology and Aging*, *30*, 279–285. https://doi.org/10.1037/pag0000024

*Zahodne, L. B., Manly, J. J., Azar, M., Brickman, A. M., & Glymour, M. M. (2016). Racial disparities in cognitive performance in mid-and late adulthood: Analyses of two cohort studies. *Journal of the American Geriatrics Society*, *64*, 959–964. https://doi.org/10.1111/jgs.14113

*Zahodne, L. B., Manly, J. J., Smith, J., Seeman, T., & Lachman, M. E. (2017). Socioeconomic, health, and psychosocial mediators of racial disparities in cognition in early, middle, and late adulthood. *Psychology and Aging*, *32*, 118–130. https://doi.org/10.1037/pag0000154

*Zahodne, L., Manly, J., Narkhede, A., Griffith, E., DeCarli, C., Schupf, N., Mayeux, R., & Brickman, A. (2015). Structural MRI predictors of late-life cognition differ across African Americans, Hispanics, and Whites. *Current Alzheimer Research*, *12*, 632–639. https://doi.org/10.2174/1567205012666150530203214

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols*, *1*, 297–301. https://doi.org/10.1038/nprot.2006.46

*Zhu, W. (2015). *Association of Objectively Measured Physical Activity with Cognitive Function in Black and White Older Adults: Reasons for Geographic and Racial Differences in Stroke (REGARDS) study* [Doctoral dissertation]. Arizona State University.

Zsembik, B. A., & Peek, M. K. (2001). Race differences in cognitive functioning among older adults. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, *56*, S266–S274. https://doi.org/10.1093/geronb/56.5.S266

Received June 8, 2020
Revision received October 13, 2020
Accepted October 14, 2020 ■