A meta-analysis of the relationship between socioeconomic status and executive function performance among children

Gwendolyn M. Lawson[a], Cayce J. Hook[b], & Martha J. Farah[a]

[a] University of Pennsylvania, 425 S. University Avenue, Philadelphia PA, 19104, United States

[b] Stanford University, Stanford, CA, United States

Corresponding author: Gwendolyn M Lawson
glawson@psych.upenn.edu
3720 Walnut Street
Philadelphia, PA 19104, United States
telephone: 402-301-1256
fax: 215-898-7301

Author note

Research Highlights

1. Socioeconomic status has often been reported to predict childhood executive function. Here the strength of this relationship across studies was assessed using meta-analysis.

2. SES and EF were significantly associated across all studies, with substantial heterogeneity in effect size.

3. When all studies were considered together, small effects were found. Studies with more SES variability in the samples and more of EF measures resulted in larger effect size estimates, in the small-to-medium range

Abstract

The relation between childhood socioeconomic status (SES) and executive function (EF) has recently attracted attention within psychology, following reports of substantial SES disparities in children's EF. Adding to the importance of this relation, EF has been proposed as a mediator of socioeconomic disparities in lifelong achievement and health. However, evidence about the relationship between childhood SES and EF is mixed, and there has been no systematic attempt to evaluate this relationship across studies. This meta-analysis systematically reviewed the literature for studies in which samples of children varying in SES were evaluated on EF, including studies with and without primary hypotheses about SES. The analysis included 8,760 children between the ages of 2 and 18 gathered from 25 independent samples. Analyses showed a small but statistically significant correlation between SES and EF across all studies (r_{random} = .16, 95% CI [.12, .21]) without correcting for attenuation due to range restriction or measurement unreliability. Substantial heterogeneity was observed between studies, and a number of factors, including the amount of SES variability in the sample and the number of EF measures used, emerged as moderators. Using only the 15 studies with meaningful SES variability in the sample, the average correlation between SES and EF was small-to-medium in size (r_{random} = .22, 95% CI [.17, .27]). Using only the 6 studies with multiple measures of EF, the relationship was medium in size (r_{random} = .28, 95% CI [.18-.37]). In sum, this meta-analysis supports the presence of SES disparities in EF and suggests that they are between small and medium in size, depending on the methods used to measure them.

Executive function (EF) refers to the cognitive processes, supported by prefrontal cortex, that regulate goal-directed behavior (Miller & Cohen, 2001). EF develops throughout childhood and adolescence, with individual differences observed from infancy (e.g., Diamond, 2001) through adulthood (e.g., Miyake & Friedman, 2012). A recently discovered predictor of such differences, documented in a growing literature within psychology and education, is childhood socioeconomic status (SES). SES refers to a combination of economic resources (e.g., income and material wealth) and social resources (e.g., social prestige and education) and correlates with a variety of family characteristics, such as parenting behavior and frequency of stressful life events (e.g., Duncan & Magnuson, 2012).

SES disparities in EF among children have been demonstrated with a wide array of tasks, including Stroop-like tasks, digit span and dimensional card sorting (e.g., Blair et al., 2011; Dilworth-Bart, 2012; Hughes & Ensor, 2005; Mezzacappa, 2004; Sarsour, Sheridan, Jutte, Nuru-Jeter, Hinshaw & Boyce, 2011). Additionally, SES disparities in EF appear larger than disparities in other cognitive abilities. In three studies comparing SES disparities across neurocognitive systems in children, disparities in EF were larger than disparities in most other neurocognitive domains (Farah et al., 2006; Noble, Norman & Farah, 2005; Noble, McCandliss & Farah, 2007). However, not all studies have found SES differences in childhood EF (e.g., Engel, Santos & Gathercole, 2008; Lupien, King, Meaney & McEwen, 2001; Wiebe, Espy & Charak, 2007).

In view of these mixed results, it is possible that SES and EF are only weakly correlated or even that they are uncorrelated, with some combination of publication bias and citation bias leading to the generalization that SES predicts EF. Alternatively, the null results may be explained by small but real correlations combined with chance error, or systematic factors such

as stringent exclusionary criteria for health and cognitive ability resulting in exceptionally healthy and able low SES subjects (Hackman, Farah & Meaney, 2010).

Understanding the relationship between SES and EF is important for at least three reasons. First, the basic science of human development involves understanding the nature of individual differences in cognition and their association with developmental contexts. Much research has examined the relation between extreme environmental adversities, such as psychosocial deprivation and abuse (e.g., Pollak et al., 2010; Hostinar, 2012) and the development of cognitive systems, particularly the prefrontal system of executive function. More recently work has begun to examine whether development of neural systems also varies with contexts within the normal range of childhood experience, such as those associated with childhood socioeconomic status. This work frequently identifies EF and its prefrontal substrates as associated with childhood SES (e.g., Kishiyama, Boyce, Jimenez, Perry, & Knight, 2008; Lawson et al., 2013; Sheridan et al., 2012).

Second, at a more practical level, executive function predicts a variety of important life outcomes, including academic achievement (e.g., Best, Miller & Naglieri, 2011), health behaviors (e.g., Williams & Thayer, 2009) and mental health (e.g., Rogers et al., 2004). These outcomes are themselves positively associated with SES. The relevance of assessing the SES-EF relation lies partly in the potential role of EF as a mediator of SES disparities in these outcomes. Indeed, a number of interventions have specifically targeted selective attention or executive function as a means to reducing SES disparities in academic achievement (e.g., Diamond & Lee, 2011; Neville et al., 2013). If the relationship between SES and EF is weak or nonexistent, it is unlikely that EF is a meaningful mediator of SES disparities in cognitive and health outcomes and such interventions would hold less promise.

Third, the relation of SES to EF is of relevance to developmental psychology research more broadly. Even for studies whose hypotheses are unrelated to SES, the SES of participants may affect results and should therefore be considered. The magnitude of the SES-EF relationship will determine how consequential unmeasured or uncontrolled SES would be in such studies.

In addition, because the extant literature on SES and EF is inconsistent, a quantitative synthesis of this literature offers the opportunity to identify study features (e.g., sample population, the measurement of SES or EF) that may help explain when and why SES disparities are found and when and why they are not. The present meta-analysis provides the first quantitative synthesis of studies reporting correlations between SES and EF.

Measuring Socioeconomic Status

The term socioeconomic status (SES) is used to refer to a family's access to economic and social resources. SES can be estimated with measures of family income, parental education level, or parental occupational prestige. Researchers sometimes combine two or more such measures to estimate overall SES. However, some have argued that components of SES – such as family income and parental education should be examined separately (Braveman et al., 2005; Duncan & Magnuson, 2012; Geyer, Hemström, Peter, & Vågerö, 2006). These researchers note that these components have different degrees of stability across time and are likely to be responsive to different policy interventions (Duncan & Magnuson, 2012). Rather than assume that all measures of SES are equally predictive of EF, or select a particular measure a priori, we take advantage of the full range of SES measures used in the EF literature by including type of SES as a moderator. We are thereby able to examine whether the measures used to estimate child SES influences the magnitude SES-EF relationship.

Here it is worth noting that SES and poverty are distinct, though related, concepts. Poverty corresponds most closely to the lowest end of the SES continuum. Although poverty is a pressing social concern, many important life outcomes including health and academic achievement show a gradient over the full range of SES (e.g., Adler et al., 1994; Reardon, 2011), and the current meta-analysis therefore measures SES continuously across the full range of family income, parental education, and parental occupational prestige.

Although SES is typically measured using the variables just mentioned, these variables need not be the proximal causes of the observed SES disparities. Indeed, it is widely assumed that some combination of factors associated with SES play causal roles, including (but not limited to) parenting practices, exposure to stressors and school or daycare quality. The proximal variables or mediators of the SES-EF relationship is an important topic for research, but it is not the focus of the present meta-analysis. Furthermore, the research summarized by this meta-analysis was not designed to identify specific causes. Therefore, the present meta-analysis is confined to answering questions about the relation of SES and EF, including the moderating effect of how SES is measured, but cannot reveal the specific causal pathways through which SES and EF are associated.

Measuring Executive Function

The measurement of EF is similarly multifactorial, related to the multifactorial nature of EF itself. One prominent framework proposes that EF is composed of three related but separable basic components: updating information in working memory, shifting attention, and inhibiting prepotent responses. According to this model, these three basic components contribute differentially to performance on complex EF tasks (Miyake et al., 2000). Although there is mixed evidence about the extent to which the structure of EF is consistent across development,

studies of EF in childhood commonly conceptualize EF tasks in terms of working memory, attention shifting, and inhibition (Best & Miller, 2010). Therefore, the current meta-analysis employed this framework to classify EF tasks, and examined correlations between SES and separate components of EF, as well as overall EF.

Goals of Current Meta-Analysis

The current meta-analysis provides a systematic, quantitative synthesis of existing research, aimed at advancing our knowledge of childhood SES disparities in EF. Specifically, it is intended to answer several key questions. First, is there a relation between SES and EF in typically developing children? Second, how strong is that relation across studies?

The third question concerns the variability among study outcomes: Is it due simply to random error, or is the literature heterogeneous, with studies truly differing in the effect sizes they are capturing? Fourth, to what can any such heterogeneity be attributed? Potential moderators, that is, factors that account for differences in effect sizes, include features of the sample (e.g., mean age, SES variability), and study design (e.g., operationalization of SES and EF). The identification of moderators may help explain disparate findings in the literature.

A special case of the moderator question concerns the hypothesis noted earlier, that different components of SES such as parental education and income may impact EF development differently. This is difficult to assess with individual studies, as few have included multiple measures of SES. However, by combining information from multiple studies, we can begin to estimate separate effect sizes for income-based, education-based, occupation-based and composite SES measures.

The current meta-analysis also advances our understanding by broadening the set of studies brought to bear on these questions about SES disparities in EF. The literature generally

cited in this connection is focused specifically on SES and EF. In contrast, a much larger

literature exists in which EF has been measured in connection with a wide range of topics, and

SES has been measured as a covariate. By meta-analyzing this larger literature, we broaden the

population of studies relevant to the topic of SES disparities in EF and also reduce the risk of

publication bias affecting our conclusions (Cooper, 2010).

Method

Search Procedures and Selection of Studies

Literature search. Relevant studies were identified through searches of the databases

PsycINFO & ERIC through January, 2013 using keywords for executive function and for

socioeconomic status. The search required that studies use at least one of the following

executive function keywords in the abstract: *executive function, cognitive control, executive*

functioning, self-regulation, working memory, inhibition, inhibitory control, shifting, cognitive

flexibility, attention, prefrontal. Identified studies also used at least one of the following

socioeconomic status keywords in the entire paper: *socioeconomic status, SES, socio-economic*

status, social status, income, poverty, disadvantaged, parental education. Unpublished

dissertations, in addition to published journal articles, were included in order to minimize the

effects of publication bias. This search identified a total of 2711 results, which were screened for

the inclusion criteria. An additional 19 potentially relevant studies were identified by reviewing

the citations of articles identified in this search and by searching the work of relevant

researchers.

Inclusion Criteria. To be included in the meta-analysis, studies were required to meet the

following inclusion criteria:

      a) Published or unpublished empirical paper

b) Was written between 1980 and January, 2013

b) Includes at least one measure of a behavioral, neurocognitive task of executive function (EF)

c) Includes at least one variable, from the following list, as a measure of socioeconomic status (SES): income, parental education, parental occupation, or some combination of these measures

d) Uses a sample of children who were between the ages of 2 and 18, at one or more time-point when data are reported

e) The population was not selected for any physical or mental disorder (e.g., ADHD, depression, low birth weight) or special condition (e.g., bilingualism, homelessness) present in the children or the parents

f) The population represented a continuous distribution of socioeconomic status

g) One or more zero-order Pearson correlations between EF and SES variables were reported in the paper or could be obtained from the corresponding author

Eligible EF and SES Measures. Executive function can be measured in a number of ways, including performance on neurocognitive tasks, self-report questionnaires, and informant-report questionnaires (Hughes, 2011). For the purpose of this meta-analysis, studies were considered eligible for inclusion in the meta-analysis only when they included at least one behavioral task measure of executive function. Eligible tasks included measures of working memory (e.g., Digit Span tasks), attention shifting (e.g., Dimensional Change Card Sort tasks), inhibition (e.g., Stroop tasks) and other tasks commonly classified as executive function (e.g., Tower of Hanoi, Continuous Performance tasks). Studies that reported only questionnaire measures of executive function or delay-of-gratification measures were not eligible for inclusion.

Similarly, a number of measures are commonly used to assess socioeconomic status. As already noted, most definitions of SES conceptualize it as a combination of family income, parental education, and parental occupation (Duncan & Magnuson, 2012). Therefore, studies were considered eligible for inclusion only when they included at least one variable that is a measure of family-level SES: family income (e.g., household income, income-to-needs ratio), parental education (e.g., maternal education, paternal education), or parental occupation (e.g., maternal occupation, paternal occupation). Composite SES measures including two or more of the aforementioned measures, including those that also included a measure of family wealth, were eligible for inclusion. Studies that reported only other measures of SES (e.g., neighborhood disadvantage, participation in free or reduced lunch, amount of time spent in poverty) or of related sociodemographic risk factors (e.g., single-parent households) were not eligible for inclusion. Also excluded were studies that mention aspects of the sample's SES, (e.g., the proportion of the sample below the poverty line) but do not identify an SES variable of interest or a covariate and thus cannot provide information about the SES-EF relation.

SES distribution. Meta-analysis requires identifying a common effect size statistic with which to compile results across studies. Which effect size statistic is appropriate depends on the hypotheses being tested by the meta-analysis and the nature of the variables being analyzed (Lipsey & Wilson, 2001). Because the majority of identified papers, including those that did not have primary hypotheses about SES, used samples with continuous SES distributions, Pearson correlations were used as the effect size measure in the present meta-analysis.

Studies comparing children from "higher SES" or "lower SES" groups, drawn from a continuous SES distribution, were included when enough information was reported or obtained

from study authors to estimate *r*-type effect sizes. Mean difference-type effect sizes were converted to *r*-type effect sizes.

In contrast, extreme group designs, in which children are enrolled based on having an SES below a relatively low SES threshold or above a different and relatively higher threshold, yield effect sizes that are not comparable to each other or to the others included here. Following recommendations that it is inappropriate to apply meta-analysis to effect size estimates based on extreme groups data (Preacher et al., 2005), we excluded such studies.

Selection of studies. A flow chart depicting the search process and exclusion of studies is shown in Figure 1. After the initial search, all of the titles and abstracts were screened to eliminate articles that, based on the title and abstract alone, did not meet inclusion criteria (e.g., not an empirical paper, published outside the relevant time period, not about the relevant constructs), resulting in 193 articles identified as potentially relevant. The full text of these potentially-relevant articles were reviewed to determine eligibility according to the following screening criteria: appropriate EF measure, appropriate SES measure, subjects within relevant age range, population not selected for any disorder or special condition, population represented continuous SES distribution. 42 articles met these inclusion criteria. To assess the reliability of this screening process, 36 articles (approximately 18.6%) were screened by two individuals, which yielded a Kappa of .83, which is considered "almost perfect" agreement (Landis & Koch, 1977).

These 42 articles were then screened to determine whether they reported one or more correlations between SES variables and EF variables in the article, or reported enough information for at least one correlation to be calculated. The corresponding authors of articles that did not report enough information to calculate an *r*-type effect size for unique samples were

contacted to request this information. Of the 42 articles that met inclusion criteria, 25 articles

reported one or more correlations between SES variables and EF variables. Additionally,

correlations were received by email for 8 articles. 9 articles were excluded because they did not

report the relevant correlations or provide them by email. Thus, 33 articles, representing 25

datasets, were eligible for inclusion in the meta-analysis.

Effect Size and Moderator Coding Procedure

All articles were coded for effect sizes and moderators using a formal coding manual, and

the first author made all final coding decisions. Additionally, a research assistant was trained on

the coding procedure and coded moderators and effect sizes for 94% of articles. Inter-rater

reliability analyses were performed to determine consistency among raters. Kappa statistics are

reported for nominal moderators. Based on Landis & Koch's (1977) guidelines, Kappa values of

.81-1.0 were considered "almost perfect agreement," and Kappa values of .61-.80 were

considered "substantial agreement." Intraclass Correlation Coefficient (ICC) statistics are

reported for effect size information (e.g., Pearson's r's and sample sizes) and continuous or

interval level moderators (e.g., mean age). Based on Fleiss's (1986) guidelines, values of ICC

above .75 were taken to represent "excellent" reliability.

Effect Size coding. Pearson correlations and samples sizes were recorded for each

correlation between SES measures and EF measures reported in each article. Pearson

correlations were reverse coded as appropriate (e.g., in cases where a higher score on the EF

variable indicates worse performance, or a higher score on the SES variable indicates lower SES)

such that a positive correlation indicated a that higher SES is associated with better EF. Sample

sizes were recorded as reported for each effect size, or were estimated as needed using reported

information about total sample size and percentage of missing data. The interrator reliability for

the raters on Pearson correlations was ICC (3,1) = .92, and the interrator reliability on sample sizes for all coded effect sizes was ICC (3,1) = .98.

Moderator coding. Following Lipsey and Wilson (2001), moderators are organized based on whether they are characteristics of the sample (e.g., sample age, gender composition) or of the measures of SES or EF used to estimate the effect size. Sample characteristics should be the same for all correlations within the sample, whereas effect size characteristics may vary for different correlations reported within the same sample (e.g., a study reporting a correlation for income-EF and a separate correlation for parental education-EF). Additionally, we include two potential moderators related to publication characteristics because they could vary between different publications from the same study.

Sample characteristics. Five characteristics of the sample were coded for each sample using information from all published and unpublished papers included in the meta-analysis. In all cases, when information about a particular sample characteristic was not reported in any publications about the sample, that sample characteristic was coded as Not Reported and the study was excluded from the appropriate moderator analysis. Additionally, in cases where multiple publications from the same study received different codes for a moderator variable, the study was excluded from the appropriate moderator analysis.

Age range. The range between the youngest and oldest participant in the study was coded in the following categories: < 2 years, 2 – 3.99 years, 4 – 6.99 years, > 7 years. The interrater reliability on this variable was Kappa = .81 ($p < .001$).

Intended sample SES. The SES distribution of the intended sample, as described by the paper(s), was coded into the following nominal categories: Low SES (e.g., studies with the stated goal of recruiting a low-SES or "at risk" sample, samples recruited from Head Start), Middle

SES (e.g., studies with the stated goal of recruiting a middle-SES sample), High SES (e.g.,

studies with the stated goal of recruiting a high-SES sample), Representative/Diverse (e.g.,

studies with the stated goal of recruiting subjects of diverse SES), Convenience Sample (e.g.,

studies with no stated goal of recruiting a particular SES range). The interrater reliability on this

variable was Kappa = .71 ($p < .001$).

   Amount of SES variability in the sample. The SES variability of the sample was coded

into two categories: Meaningful Variability Reported and Meaningful Variability Not Reported.

We were not able to use a specific threshold to make this coding decision because the

information that papers reported regarding sample variability was not consistent across papers.

Instead, studies were categorized as 'Meaningful Variability Reported' when the paper described

the sample as heterogeneous or reported a substantial amount of SES variability in the sample,

and were categorized as 'Meaningful Variability Not Reported' when the paper described the

sample as homogenous (e.g., "a sample of middle-socioeconomic status kindergartners" as in

Cameron et al., 2012), reported a small amount of SES variability in the sample (e.g., a sample in

which the standard deviation for family income is under $7,000 and only 14.2% of caregivers are

classified as having an associates or bachelors degree as in Rhoades, Greenberg & Domitrovich,

2009) or did not include a description of the SES variability of the sample. The interrater

reliability on this variable was Kappa = .75 ($p = .001$). The information used to make this coding

decision for each independent sample is displayed in Table 1.

   Extent of exclusionary criteria. Studies were also coded for the extent to which the

sample selection used exclusionary criteria based on physical health, mental health, and/or

cognitive ability that required participants to be healthy or high functioning. Stringent criteria

would be expected to attenuate an SES effect (Hackman, Farah & Meaney, 2010). They were

coded into two categories: Minimal (e.g., no exclusionary criteria based on health or ability, or criteria that excluded only extreme cases, such as children with mental retardation), and High (e.g., one or more exclusionary criterion based on health or ability, excluding more than extreme cases). The interrater reliability for the raters on this variable was Kappa = .60 ($p < .001$).

Racial/ethnic composition. The racial/ethnic composition of the sample was coded categorically, based on which racial/ethnic group (e.g., White, Black, Hispanic, Other) predominated (i.e., composed greater than 60% of the sample), or whether the sample was mixed (such that no group composed greater than 60% of the sample). If this information was not reported, the sample was coded as Not Reported.

Mean age. The mean age (in years) of the sample at the time when the EF variable(s) used to obtain effect sizes were collected was coded. When effect sizes from multiple time-points were reported, the mean ages at these time-points were averaged. The interrater reliability for the coders on this variable was found to be ICC $(3,1) = .95$ ($p < .001$).

Sex composition. Studies were coded for the percentage of the sample that was male. When multiple publications from the same sample had different sex compositions, the average was used in analyses. The interrater reliability for the coders on this variable was found to be ICC $(3,1) = 1$ ($p < .001$).

Effect size characteristics. Four effect size characteristics were coded for each reported correlation between an SES variable an EF variable.

Category of SES construct. For each effect size, the category of the SES construct was coded into the following nominal categories: income-based constructs (e.g., family income, income-to-needs ratio), education-based constructs (e.g., maternal education, paternal education), occupation-based constructs (e.g., maternal occupation, paternal occupation), composite

constructs (e.g., constructs using two or three of the aforementioned constructs). The interrater reliability on this variable was Kappa = 1.0 ($p < .001$).

Number of measures used to calculate the SES variable. For each effect size, the number of measures used to obtain the SES variable was coded. The interrater reliability for the coders on this variable was found to be ICC (3,1) = .98 ($p < .001$).

Category of EF construct. For each effect size, the category of the EF construct was coded in the following categories: working memory, attention shifting, inhibition, composite or latent variables using 2 or 3 of the aforementioned categories, or other (e.g., planning). The interrater reliability on this variable was Kappa = .95 ($p < .001$).

Number of measures used to calculate the EF variable. For each effect size, the number of measures used to obtain the EF variable was coded. The interrater reliability for the coders on this variable was found to be ICC (3,1) = .85 ($p < .001$).

Publication characteristics. Two publication characteristics were coded for each publication.

Type of publication. The article was coded as either a published journal article or an unpublished thesis or doctoral dissertation. By definition, publication bias will influence results obtained from published, but not unpublished, studies. The interrater reliability for the raters on this variable was Kappa = 1.0 ($p < .001$).

SES as a primary focus. Some studies were undertaken with an interest in the SES-EF relation, whereas others used SES as a covariate in a study of EF. One would expect SES effects to be larger in the first case, in part because such studies might invest more care in the measurement of SES and in addition because such studies would be subject to a bias against publishing small or nonexistent effects of SES. To examine this possibility, the focus of each

paper was coded into two categories: SES as a Primary Focus and SES Not as a Primary Focus.

Papers were classified as 'SES as a Primary Focus' when hypotheses or study goals pertaining to

SES were stated in the abstract or introduction of the paper. Papers were classified as 'SES Not

as a Primary Focus' when hypotheses or study goals about SES were not stated in the abstract or

introduction of the paper.

The interrater reliability for the raters on this variable was Kappa = .75 ($p < .001$).

Analytical Procedures.

Transformations, calculations of weighed mean effect sizes, tests for heterogeneity, and

moderator analyses were conducted in Comprehensive Meta-Analysis (V.3.3.070, November

2014, Biostat, Englewood-USA).

Calculating average effect sizes. The Pearson product-moment correlation coefficient r

was the primary effect size measure used in analyses. Because the product-moment correlation

has a problematic standard error formulation in its standard form (Alexander, Scozzaro, &

Borodkin, 1989), the correlations were transformed using Fisher's Z_r-transform (Hedges &

Olkin, 1985.

Following the recommendations of Hedges & Olkin (1985), each effect size was then

weighted by its inverse variance weight in order to account for its precision. This weighting

procedure gives greater weight to larger samples than smaller samples (Lipsey & Wilson, 2001).

For ease of interpretation, Fisher's Z_r-transformed correlations were transformed back into the

standard correlational form for the presentation of results.

Statistical independence. Meta-analysis assumes that observations used in analyses are

independent of each other. Several steps were taken in order to meet this assumption. First,

datasets, rather than publications, were used as the unit of analysis. In cases where multiple

articles represented the same datasets (e.g., multiple dissertations and published papers using data from the Family Life Project), effect sizes were averaged across all reported effect sizes from all included articles to generate one effect size per dataset. Papers with partially overlapping samples were also treated as the same dataset. Additionally, many studies included in the meta-analysis reported multiple correlation coefficients based on a single sample. In order to avoid violations of statistical independence, multiple correlations per dataset were averaged, such that each dataset only contributed one effect size to the calculation of the average effect size and to tests of moderation by study and publication characteristics.

Fixed and random effects models. There is ongoing disagreement about whether it is more appropriate to use fixed or random effects model in meta-analysis (Lipsey & Wilson, 2001). Fixed effects models assume that there is a common true effect size across studies, with random error that stems only from subject-level sampling error in each study. In contrast, random effects models assume that the true effect size varies across studies, due to systematic variability across studies, in addition to subject-level sampling error. Unlike fixed effects models, random effects models allow the results to be generalized to studies not included in the analysis (Borenstein et al., 2009). Because meaningful variation in effect sizes between studies was anticipated, random effects models were deemed most appropriate for the overall analysis and mixed effects were considered most appropriate for moderator analyses. For the sake of comparison, results of the overall analysis and moderator analyses are also reported using fixed effects models.

Tests for heterogeneity. The heterogeneity among effect sizes was examined using the Q statistic and the I^2 statistic. The Q statistic provides a significance test indicating whether the observed range of effect sizes is larger than would be expected from within-study variance alone

However, the Q statistic has low power to detect true heterogeneity, especially in situations where a small number of studies are included in the meta-analysis (Hardy & Thompson, 1998). Therefore, the I^2 index was also used to examine heterogeneity among studies. The I^2 index ranges from 0 to 100 and describes the percentage of total variance that is attributed to between-study variance. I^2 values of 25%, 50%, and 75% have been used as benchmarks representing low, moderate, and high heterogeneity, respectively (Higgins, Thompson, Deeks, & Altman, 2003).

Moderator analyses. Mean sample age, number of SES measures, and number of EF measures were assessed using meta-regression with random effects. All other moderators were were categorical and thus assessed using an analysis of variance (ANOVA) of mixed-effects models for each potential moderator. For the sake of comparison, moderators were also assessed with the Q_b test using fixed-effects models.

It is worth noting that many studies reported effect sizes representing multiple levels of a given effect size characteristic (e.g., EF-family income and EF-parental education correlations reported by a single study). This required consideration when testing for moderation by categorical effect size characteristics (e.g., category of SES construct, category of EF construct). The primary analyses used a "shifting units of analysis" approach (Cooper, 2010), which allowed each study to contribute one effect size per level of the effect size characteristic being tested in a given moderation analysis. This approach slightly relaxes assumptions of independence, but allows more data to be utilized in tests of moderation. Studies for which moderator variables could not be coded from the information provided in the study and subgroups including only one study were excluded from moderator analyses.

Tests for publication bias. Several methods were used to assess for publication bias. First, we created funnel plots. The effect size for each dataset was plotted against the study precision (inverse of standard error). The lower the precision of studies, the greater the dispersion of effect sizes around the true value, making the shape of the scatterplot like an upside down funnel. If publication bias has caused nonsignificant small effects to go unreported, then the funnel plot will be negatively skewed, with missing points in the lower left part of the plot (Sterne, Becker & Egger, 2005). In addition, Duval and Tweedie's (2000) trim-and-fill procedure was used to impute missing studies and compute the summary effect size correcting for the number and assumed location of the missing studies. The classic fail-safe N value was also computed to determine the number of missing studies that would bring the p value above .05, and this value was compared to Rosenthal's tolerance level for an unlikely number of nonsignificant studies (computed as $5K + 10$, where K is the number of observed studies; Rosenthal, 1979). Additionally, Orwin's fail-safe N value (Orwin, 1983) was computed to determine the number of missing studies that would bring the Fisher's Z score to a trivial effect size. It is important to note that funnel plots and the trim-and-fill procedure rely on the assumption of homogeneity of effect sizes (Terrin et al., 2003). Therefore, results from these techniques should be interpreted with caution in heterogeneous data sets.

Results

Study characteristics. Table 2 displays characteristics of the papers used in this analysis. There were 33 papers representing 25 independent samples. 18 of the 25 studies took place in the United States; other studies took place in Canada, Germany, the Netherlands, Hong Kong, Turkey, and Madagascar. A total of 111 effect sizes were coded from these studies. Individual correlations between SES variables and EF variables ranged from -.11 to .48. Table 3 displays

average effect size information for each independent sample, with average correlations ranging

from -.04 to .47. According to convention (Cohen, 1988) correlations of .1, .3 and .5 are

considered small, medium and large, respectively.

Overall effect size. Taking all studies together, the strength of relation between children's SES

and EF was $r = .16$, 95% CI [.12, .21] using the random effects model, and $r = .18$, 95% CI [.16,

.20] using the fixed effects model. These are conventionally considered "small" effect sizes.

They were, however, significantly different from zero ($z = 6.55$, $p < .001$ for random effects

model, $z = 15.20$, $p < .001$ for fixed effects model). The forest plot for these analyses is shown

in Figure 2.

Tests for heterogeneity. The Q statistic indicated significant heterogeneity among the studies,

$Q (24) = 92.89$, $p < .001$, as did the I^2 index of 74.16 (Higgins et al., 2003). In order to test

hypotheses about why some studies showed larger effect sizes than others, moderator analyses

were performed.

Moderator analyses.

Sample characteristics. Eight characteristics of the samples were assessed as

moderators: intended sample SES, amount of SES variability, extent of exclusionary criteria for

the sample, racial/ethnic composition of the sample, age range, mean age, and sex composition

of the sample. Results of moderator analyses for categorical sample characteristics are displayed

in Table 4. Of these, only amount of SES variability ($Q (1) = 14.79$, $p < .001$) emerged as a

significant moderator using mixed effects models. Studies with meaningful SES variability ($r =$

.22; $k = 15$) had significantly larger SES effect sizes than studies without meaningful SES

variability ($r = .08$; $k = 9$). Using fixed effects models, racial composition ($Q_b (2) = 13.82$, $p =$

.001) and age range ($Q_b (2) = 15.20$, $p < .001$) also emerged as significant moderators, and SES

variability remained a significant moderator (Q_b (1) = 35.58, $p < .001$). Studies using samples that were greater than 60% Black ($r = .06$; $k = 2$) had smaller effect sizes than studies with samples that were greater than 60% White ($r = .16$; $k = 7$) or were mixed race/ethnicity ($r = .22$; $k = 10$). Studies with a sample age range of < 2 years ($r = .21$; $k = 13$) had larger effect sizes than studies with sample ages ranges of 2-3.99 years ($r = .12$; $k = 10$). Meta-regression with random effects was used to assess mean age of the sample and sex composition of the sample as potential moderators. Neither mean age (slope = -.0008, $p = .93$) nor sex composition of the sample (slope = -.003, $p = .58$) were significantly associated with effect size, providing no evidence that either of these sample characteristics were associated with effect size.

Effect size characteristics. Four effect size characteristics related to the measurement of EF and SES were assessed as moderators: category of the SES construct, category of the EF construct, number of SES measures, and number of EF measures. As previously noted, several studies reported effect sizes for multiple levels of these effect size characteristics. Specifically, 9 studies reported effect sizes for two or more categories of SES construct, and 13 studies reported effect sizes for two or more categories of EF construct. Additionally, 2 studies reported correlations for two or more levels of the "number of SES measures" moderator, and 6 studies reported correlations for two or more levels of the "number of EF measures" moderator.

Meta-regression with random effects was used to assess "number of SES measures" and "number of EF measures" as moderators. For studies reporting effect sizes for more than one level of these variables, the effect size for the highest number of measures reported by the study was used in the meta-regression. As previously noted, the primary tests of moderation by category of SES construct and category of EF construct were conducted using a "shifting units of

analysis" approach (Cooper, 2010), in which a given study was allowed to contribute one ES per level of the effect size characteristic being tested as a moderator.

The meta-regression results revealed that the number of EF measures was significantly associated with effect size (slope = .06, $p < .001$). A scatterplot of the relationship between number of EF measures and effect size is shown in Figure 3. The relationship between the number of SES measures and effect size was marginally-significant (slope = .06, $p = .08$), and a scatterplot of the relationship between number of SES measures and effect size is shown in Figure 4.

Results of moderator analyses for effect size characteristics using a "shifting units of analysis" approach are displayed in Table 5. Using mixed effects models, category of EF construct (Q (4) = 10.66, $p = .03$) significantly moderated effect sizes. The average correlation for effect sizes using composite EF measures was $r = .28$ (95% CI [.18-.37]), as compared to $r = .18$ (95% CI [.13-.22]) for working memory, $r = .17$ (95% CI [.08-.25]) for attention shifting, $r = .14$ (95% CI [.07-.22]) for inhibition, and $r = .09$ (95% CI [.06-.16]) for other measures of EF. Category of SES construct (($Q(2) = .79$, $p = .67$) did not significantly moderate effect sizes.

Publication characteristics. Two characteristics of the publications were assessed as moderators: type of publication and whether or not the publication focused on SES. Studies, rather than publications, were used as the unit of analysis in these analyses. Studies were excluded from analysis in cases where multiple publications from the same study received different codes on one of these characteristics (e.g., a published paper and an unpublished dissertation from the same study).

Results of moderator analyses for publication characteristics are displayed in Table 6. Using mixed effects models, type of publication (Q (1) = 5.05, $p = .03$) significantly moderated

effect sizes, with larger effect sizes for published ($r = .18$; $k = 18$), as compared to unpublished ($r = .08$; $k = 5$), papers.

Publication bias.   In order to determine the extent to which publication bias may have impacted the results of the current meta-analysis, publication bias was first assessed using all articles included in the meta-analysis. Standard errors for all datasets were plotted against effect size to generate a funnel plot. This plot is shown in Figure 5. In the absence of publication bias, studies should be distributed symmetrically around the average effect size. If publication bias is present, the bottom of the plot tends to show a higher concentration of studies to the right of the mean. Duval & Tweedie's (2000) trim-and-fill procedure was used to correct for missing studies. Using the random effects model to look for missing studies to the left and right of the mean, 0 missing studies to the left of the mean were identified. The classic fail-safe N value indicates the number of non-significant studies that would be needed to nullify the effect (Rosenthal, 1979). The classic fail-safe N value was 1111, well above Rosenthal's tolerance level of 135 for an unlikely number of nonsignificant studies (Rosenthal, 1979). Orwin's fail-safe N indicates the number of studies with an effect size of zero that would be needed to bring the aggregated correlation to a 'trivial' size. Using $r = .05$ as the criterion for a trivial correlation, Orwin's fail-safe N value was 65.

Discussion

The current paper provides the first meta-analytic synthesis of the literature on the relationship between socioeconomic status and executive function performance among children. SES was significantly associated with EF, although the strength of the association varied markedly between studies. We were able to identify several factors that influenced the size of the SES-EF relationship.

Across the 25 studies included in the meta-analysis, an overall correlation or $r = .16$ between SES and EF was observed. However, a number of studies included in the meta-analysis were not designed to answer questions about SES and little SES variability in their samples, and corrections for range restriction and other artifacts were not made. Of the 15 studies with meaningful SES variability, the overall correlation was $r = .22$, suggestive of a small-to-medium relationship between SES and EF among socioeconomically diverse populations. The results of tests for publication bias did not suggest that these results were inflated by publication bias.

A primary objective of this meta-analysis was to investigate factors that moderate the strength of the relationship between SES and EF. This aim was particularly important given the large amount of heterogeneity between studies that was observed (Borenstein et al., 2009). The amount of SES variability in the sample emerged as a significant moderator, with meaningful SES variability associated with larger effect sizes. This conclusion must be qualified by the lack of objective and comparable criteria available across studies because studies varied greatly in the SES variability information they reported (e.g., means and standard deviations). Nevertheless, this moderator was coded with good inter-rater reliability, and it is likely that error in measuring this factor would decrease, rather than increase, its association with effect sizes. Further, this result is not surprising considering the fact that range restriction attenuates correlations (Hunter & Schmidt, 2004). These results therefore suggest that differing amounts of SES variability in samples may be a factor explaining discrepancies in observed relationships between SES and EF between studies. Thus, it is important for researchers to clearly report the amount of SES variability in samples (e.g., means and standard deviations of SES variables) and to consider range restriction when interpreting results.

Additionally, the way in which EF was measured influenced the magnitude SES-EF relationship. In particular, a higher number of measures used to calculate the EF variable related to a larger SES-EF effect size. Category of EF construct also emerged as a significant moderator, with composite EF measures showing the largest effect sizes (r = .28). One potential explanation for these results is reduced measurement error in EF variables derived from multiple measures (Spearman, 1910). This is particularly relevant in light of recent efforts to develop EF task measures with acceptable psychometric properties in response to the observation that most EF tasks commonly used with children have not undergone psychometric evaluation (e.g., Beck et al., 2011; Willoughby et al., 2010). Furthermore, when test-retest reliability of single EF measures in children has been reported, it has been found to be low by psychometric criteria (e.g., Bishop et al., 2001), which would attenuate correlations with other variables, such as SES.

To discover whether the relationship between SES and EF widens or narrows across time, we examined the mean age of the sample as a moderator. We found no significant relationship between this factor and the size of the SES-EF relationship. This is consistent with the small number of longitudinal studies that have examined the trajectory of the SES-EF relationship (Hackman et al., 2014; Hackman et al., 2015; Hughes, Ensor, Wilson, & Graham, 2010). Collectively, the present meta-analysis and the aforementioned studies suggest that the SES-EF relationship remains stable across childhood, rather than growing with the accumulation of SES-related experiences or narrowing as low-SES children "catch up" from a developmental delay.

While mixed-effect models did not find moderation by racial composition of the sample, fixed effects models did, with predominantly Black samples showing smaller effect sizes. However, these results should be interpreted cautiously, as only two samples were classified as greater than 60% Black, and both samples were also predominantly low SES, without

meaningful SES variability reported. In addition to addressing this issue by moderator analysis, we also examined the individual meta-analyzed papers for information about the interaction between race/ethnicity and SES in predicting EF. One paper reported that this interaction was not significant (Sarsour, 2007) and the other reported that the income-EF association was greater for African-American children without having tested a statistical interaction (Dilworth-Bart, Khurshid, & Lowe Vandell, 2007). In sum, neither the results of the moderator analysis nor these individual studies provide clear support for the role of race in moderating the SES-EF relation.

Meta-analysis combines studies that may vary substantially in their measurement of variables and other methodological features, and this can be a strength or a weakness. It has been argued that, by averaging results from "apples and oranges," meta-analysis may yield meaningless figures (Hunt, 1997). However, as Rosenthal and DiMatteo (2001) note, combining apples and oranges can yield especially useful information if one wants to generalize about fruit. Because the present meta-analysis combined a number of measures of SES (e.g., parental education, family income, composite measures), and a number of measures of EF (e.g., Digit Span tasks, Stroop tasks, Continuous Performance tests), it is subject to the apples and oranges criticism, but also enables is to report on the association between SES and the overall construct of EF, as well as with more specific categories (e.g., working memory, inhibition). Indeed, the moderator analyses allowed us to assess differences in the SES-EF relation depending on SES category, EF category, age, race and other potentially relevant variables.

Additionally, while the inclusion of a broad set of studies could be considered a strength of the present meta-analysis, it could also be considered a weakness in that many of the studies were not expressly designed to answer questions about the relationship between SES and EF

(e.g., studies with extremely small variance in SES) and might therefore underestimate the size of the effect. Furthermore, the meta-analysis did not make corrections for restricted range or other study artifacts. The results, therefore, should not be interpreted as the relationship between SES and EF in the population as a whole, but rather as a summary of the correlations between SES and EF available in the current literature, including studies that were not designed to detect relationships between SES and EF.

It is also important to note that the current meta-analysis examined the raw association between SES and EF, without adjusting for related constructs, such as IQ. This reflects the state of the literature; none of the studies included in the meta-analysis report estimates of the SES-EF association after controlling for IQ. There is substantial construct overlap between EF and IQ, such that it may not be not meaningful to examine EF adjusted for IQ; this is likely to be among the reasons that researchers have not controlled for IQ in the studies reviewed here (Dennis et al., 2009). However, three of the papers reported information about the SES-EF association after controlling for a measure of verbal or language ability: one found that the SES-EF association persisted (Dilworth-Bart, 2012), another found that it did not (Turner, 2010), and a third found that some measures of language ability accounted for some measures of EF (Noble et al, 2007). Of course, the finding of mediators for the SES-EF relation does not necessarily diminish the reality or the importance of that relation.

The SES-EF relation is consistent with environmental influence on EF. Stress (e.g., Evans, 2004), parenting behavior (e.g., Evans, Boxhill & Pinkava, 2008), cognitive stimulation (e.g., Bradley, Corwyn, McAdoo & Coll, 2001) and language exposure (Hoff, 2003) all vary with SES, and could contribute to the differences summarized here. It is also true that executive functioning is highly heritable in studies using behavioral genetics methods (Engelhardt et al.,

2015). EF ability likely involves interactions among numerous genetic and environmental factors (e.g., Deater-Deckard, 2014). The current meta-analysis assesses the strength of the SES-EF relation but is not able to reveal its causes.

Given the observed association between SES and EF, it is important for future research to investigate whether and how EF contributes to SES disparities in broader life outcomes. There is a plausible role for EF in shaping health behaviors, mental health, and academic achievement. The modest correlations observed between SES and EF makes it is unlikely that EF fully mediates SES disparities in academic achievement or health. However, small differences in childhood EF may have cumulative consequences across domains of development, a phenomenon that has been termed "developmental cascades" (e.g., Masten & Cicchetti, 2010). Consistent with this, EF in childhood is predictive of a wide range of outcomes (e.g., Best, Miller & Naglieri, 2011; Williams & Thayer, 2009), suggesting that small differences in EF may shape development in meaningful ways. Thus, the role of EF, as well as other neurocognitive systems, as mediators of SES disparities in achievement and health is an important topic for further investigation.

References

Note: Studies included in the meta-analyses are marked with an *.

Adler, N. E., Boyce, T., Chesney, M. A., Cohen, S., Folkman, S., Kahn, R. L., & Syme, S. L. (1994). Socioeconomic status and health: The challenge of the gradient. American Psychologist, 49(1), 15–24. doi:10.1037/0003-066x.49.1.15

Alexander, R. A., Scozzaro, M. J., & Borodkin, L. J. (1989). Statistical and empricial examination of the chi-square test for homogeneity of correlations in meta-analysis. *Psychological Bulletin, 106,* 329-331.

*Bernier, A., Carlson, S. M., Deschênes, M., Matte-Gangé, C. (2012). Social factors in the development of early executive functioning: a closer look at the caregiving environment. *Developmental Science, 15*, 12-14. doi: doi:10.1111/j.1467-7687.2011.01093.x

*Berry, D., Blair, C., Willoughby, M., Granger, D., & The Family Life Project Key Investigators. (2012). Salivary alpha-amylase and cortisol in infancy and toddlerhood: Direct and indirect relations with executive functioning and academic ability in childhood. *Psychoneuroendocrinology, 37,* 1700-1711. doi: 10.1016/j.psyneuen.2012.03.002

Beck, D. M., Schaefer, C., Pang, K., & Carlson, S. M. (2011). Executive function in preschool children: Test-retest reliability. *Journal of Cognitive Development, 12,* 169–193. doi:10.1080/15248372.2011.563485.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, *81*(6), 1641-1660. doi: 10.1111/j.1467-8624.2010.01499.x

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and

   academic achievement from ages 5 to 17 in a large, representative national sample.

   Learning and Individual Differences, 21, 327-336. doi:10.1016/j.lindif.2011.01.007

Bishop, D. V. M., Aamodt-Leeper, G., Creswell, C., McGurk, R., & Skuse, D. H. (2001).

   Individual differences in cognitive planning on the Tower of Hanoi Task:

   Neuropsychological maturity or measurement error? *Journal of Child Psychology and*

   Psychiatry, 42, 551-556.

*Blair, C., Granger, D. A., Willoughby, M., Mills-Koonce, R., Cox, M., … the FLP

   Investigators. (2011). Salivary cortisol mediates effects of poverty and parenting on

   executive functions in early childhood. *Child Development, 82,* 1970-1984. doi:

   10.1111/j.1467-8624.2011.01643.x

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). A basic introduction

   to fixed-effect and random-effects models for meta-analysis. *Research Synthesis*

   Methods, 1, 97-111. doi: 10.1002/jrsm.12

Braveman, P.A., Cubbin, C., Egerter, S., Chideya, S., Marchi, K.S., Metzler, M., & Posner, S.

   (2005). Socioeconomic status in health research: one size does not fit all. *Journal of the*

   American Medical Association, 294, 2879-2888. doi:

   http://dx.doi.org/10.1001/jama.294.22.2879

Bryck, R. L., & Fisher, P. A. (2012). Training the brain: Practical applications of neural

   plasticity from the intersection of cognitive neuroscience, developmental psychology, and

   prevention science. American Psychologist, 67(2), 87–100. doi:10.1037/a0024657

*Cameron, C., Brock, L. L., Murrah, W. M., Bell, L. H., Worzalla, S. L., Grissmer, D., &

   Morrison, F. J. (2012). Fine motor skills and executive function both contribute to

kindergarten achievement. *Child Development, 83,* 1229-1244. doi: 10.1111/j.1467-

8624.2012.01768.x

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ:

Erlbaum.

Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.).

Thousand Oaks, CA: Sage Publications.

Deater-Deckard, K. (2014). Family matters intergenerational and interpersonal processes of

executive function and attentive behavior. *Current Directions in Psychological Science*,

23(3), 230-236.

*De Jong, P. F. (1993). The relationship between students' behaviour at home and attention and

achievement in elementary school. *British Journal of Educational Psychology, 63,* 201-

213.

*Deng, M. (2008). *A developmental model of effortful control: the role of negative emotionality

and reactivity, sustained attention in mother-child interaction, and maternal sensitivity.*

(Unpublished doctoral dissertation). University of North Carolina at Chapel Hill, Chapel

Hill, NC.

Dennis, M., Francis, D. J., Cirino, P. T., Schachar, R., Barnes, M. A., & Fletcher, J. M. (2009).

Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders. *Journal

of the International Neuropsychological Society*, *15*(03), 331.

http://doi.org/10.1017/S1355617709090481

*Deprince, A.P., Winzierl, K.M., & Combs, M.D. (2009). Executive function performance and

trauma exposure in a community sample of children. *Child Abuse and Neglect, 33*, 353-

361. doi: http://dx.doi.org/10.1016/j.chiabu.2008.08.002

Diamond, A. (2001). Looking closely at infants' performance and experimental procedures in the A-not-B task. *Behavioral and Brain Sciences,* 24, 38-41.

Diamond, A. & Lee, K. (2011). Interventions shown to aid executive function development in children 4-12 years old. *Science, 333,* 959-964. doi: 10.1126/science.1204529.

*Dilworth-Bart, J.E., (2012). Does executive function mediate SES and home quality associations with academic readiness? *Early Childhood Research Quarterly, 27*, 416-425. doi: http://dx.doi.org/10.1016/j.ecresq.2012.02.002

*Dilworth-Bart, J.E., Khurshid, A., Lowe Vandell, D., (2007). Do maternal stress and home environment mediate the relation between early income-to-need and 54-months attentional abilities? *Infant and Child Development, 16,* 525-552. doi: 10.1002/icd.528

*Doan, S.N. & Evans, G.W. (2011). Maternal responsiveness moderates the relationship between allostatic load and working memory. *Development and Psychopathology, 23*, 873-880. doi: http://dx.doi.org/10.1017/S0954579411000368

Duncan, G. J. & Magnuson, K. (2012). Socioeconomic status and cognitive functioning: moving from correlation to causation. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3, 377-386. doi: 10.1002/wcs.1176

Duval, S. & Tweedie, R. (2000). Trim and fill: funnel-plot-based method of testing bias in meta-analysis. *Biometrics, 56,* 455-463. doi: 10.1111/j.0006-341X.2000.00455.x

Engel, P. M. J., Santos, F. H., & Gathercole, S. E. (2008). Are working memory measures free of socioeconomic influence? *Journal of Speech, Language, and Hearing Research*, *51*(6), 1580-1587. doi: 10.1044/1092-4388

Engelhardt, L. E., Briley, D. A., Mann, F. D., Harden, K. P., & Tucker-Drob, E. M. (2015). Genes Unite Executive Functions in Childhood. Psychological Science, 26(8), 1151–1163. doi:10.1177/0956797615577209

Farah, M. J., Shera, D. M., Savage, J. H., Betancourt, L., Giannetta, J. M., Brodsky, N. L., Malmud, E. K., & Hurt, H. (2006). Childhood poverty: specific associations with neurocognitive development. *Brain Research*, 1110, 166-174. doi:10.1016/j.brainres.2006.06.072

*Fernald, L. Weber, A., Galasso, E., & Ratsifandrihamanana, L.. (2011). Socioeconomic gradients and child development in a very low income population: Evidence from Madagascar. *Developmental Science, 14,* 832-847. doi: 10.1111/j.1467-7687.2010.01032.x

Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. New York: John Wiley.

Geyer, S., Hemström, O., Peter, R., & Vågerö, D. (2006). Education, income, and occupational class cannot be used interchangeably in social epidemiology. Empirical evidence against a common practice. *Journal of Epidemiology and Community Health*, 60(9), 804-810. doi: http://dx.doi.org/10.1136/jech.2005.041319

Hardy, R. J., & Thompson, S. G. (1998). Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine,* 17, 841-856. doi: 10.1002/(SICI)1097-0258(19980430)17:8<841::AID-SIM781>3.0.CO;2-D

*Hackman, D.A. (2012). *Socioeconomic status and the development of executive function and stress reactivity: the specific roles of parental nurturance and the home environment.* (Unpublished doctoral dissertation). University of Pennsylvania, Philadelphia, PA.

Hackman, D. A., Betancourt, L. M., Gallop, R., Romer, D., Brodsky, N. L., Hurt, H. & Farah, M. J. (2014). Mapping the trajectory of socioeconomic disparity in working memory: Parental and neighborhood factors. *Child Development, 85,* 1433-1445.

Hackman, D. A., Farah, M. J., & Meaney, M. J. (2010). Socioeconomic status and the brain: mechanistic insights from human and animal research. *Nature Reviews Neuroscience*, *11*, 651-659. doi: 10.1038/nrn2897

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

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

*Henning, A., Spinath, F.M., Aschersleben, G. (2010). The link between preschoolers' executive function and theory of mind and the role of epistemic states. *Journal of Experimental Child Psychology. 108*, 513-531. doi: http://dx.doi.org/10.1016/j.jecp.2010.10.006

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analysis. *British Medical Journal, 327,* 557–560. doi: http://dx.doi.org/10.1136/bmj.327.7414.557

Hostinar, C. E., Stellern, S. A., Schaefer, C., Carlson, S. M., & Gunnar, M. R. (2012). Associations between early life adversity and executive function in children adopted internationally from orphanages. *Proceedings of the National Academy of Sciences*, *109*(Supplement 2), 17208-17212.

Hughes, C. (2011). Changes and challenges in 20 years of research into the development of executive functions. *Infant and Child Development*, *271*, 251-271. doi: 10.1002/icd.736

Hughes, C. & Ensor, R. (2005). Executive function and theory of mind in 2 year olds: A family affair? *Developmental Neuropsychology, 28,* 645-668.

Hughes, C., Ensor, R., Wilson, A., & Graham, A. (2010). Tracking executive function across the transition to school: A latent variable approach. *Developmental Neuropsychology*, 35, 20-36. doi: 10.1080/87565640903325691

Hunt, M. M. (1997). How science takes stock: The story of meta-analysis. New York, NY: Russell Sage Foundation.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage Publications.

*Ivrendi, A. (2011). Influence of self-regulation on the development of children's number sense. *Early Childhood Education Journal, 39,* 239-247. doi: 10.1007/s10643-011-0462-0

*Jacobson, L., (2008). *The role of executive function skills in children's competent adjustment to sixth grade: Do elementary classroom experiences make a difference?* (Unpublished doctoral dissertation). University of Virginia, Charlottesville, VA.

*Kegel, C.A. & Bus, A.G. (2012). Online tutoring as a pivotal quality of web-based early literacy programs. *Journal of Educational Psychology, 104,* 182-192. doi: http://dx.doi.org/10.1037/a0025849

Kishiyama, M. M., Boyce, W. T., Jimenez, A. M., Perry, L. M., & Knight, R. T. (2008). Socioeconomic disparities affect prefrontal function in children. *Journal of Cognitive Neuroscience*, *21*, 1106-1115.

*Knipe, H., (2009). *Cognitive reasoning skills and classroom settings: a multi-level examination of student and teacher factors in math achievement.* (Unpublished doctoral dissertation). The Pennsylvania State University, State College, PA.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159-174. doi: http://dx.doi.org/10.2307/2529310

Lawson, G. M., Duda, J. T., Avants, B. B., Wu, J., & Farah, M. J. (2013). Associations between children's socioeconomic status and prefrontal cortical thickness. *Developmental Science*, *16*, 641-652.

*Li-Grining, C. P. (2005). *Social foundations of early school success among low-income children: The role of self-regulation and home, classroom, and policy contexts.* (Unpublished doctoral dissertation). Northwestern University, Evanston, IL.

Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *The ANNALS of the American Academy of Political and Social Science, 587,* 69- 81. doi: 10.1177/0002716202250791

Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.

Lupien, S. J., King, S., Meaney, M. J., & McEwen, B. S. (2001). Can poverty get under your skin? Basal cortisol levels and cognitive function in children from low and high socioeconomic status. *Development and Psychopathology*, *13*(3), 653-676. doi: http://dx.doi.org/10.1017/S0954579401003133

Masten, A. S. & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology, 22,* 491-495. doi:10.1017/S0954579410000222

*Matte-Gangé, C. & Bernier, A. (2011). Propective relations between maternal autonomy support and child executive functioning: Investigating the mediating role of child language ability. *Journal of Experimental Child Psychology, 110*, 611-625. doi: http://dx.doi.org/10.1016/j.jecp.2011.06.006

*McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology, 43,* 947-959. doi: http://dx.doi.org/10.1037/0012-1649.43.4.947

*Mezzacappa, E. (2004). Alerting, orienting, and executive attention: Developmental properties and sociodemographic correlates in an epidemiological sample of young, urban children. *Child Development, 75,* 1373-1386. doi: 10.1111/j.1467-8624.2004.00746.x

*Mezzacappa, E., Buckner, J.C., & Earls, F. (2011). Prenatal cigarette exposure and infant learning stimulation as predictors of cognitive control in childhood. *Developmental Science, 14,* 881-891. doi: 10.1111/j.1467-7687.2011.01038.x

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Reviews of Neuroscience*, *24*, 167-202.

Miyake, A. & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions of Psychological Science, 21,* 8-14. doi: 10.1177/0963721411429458.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cognitive Psychology*, *41*(1), 49-100. doi: http://dx.doi.org/10.1006/cogp.1999.0734.

Neville, H. J., Stevens, C., Pakulak, E., Bell, T. A., Fanning, J., Klein, S., & Isbell, E. (2013). Family-based training program improves brain function, cognition, and behavior in lower socioeconomic status preschoolers. *Proceedings of the National Academy of Sciences, 110(*29), 12138-12143. doi: 10.1073/pnas.1304437110.

*NICHD Early Child Care Research Network (2005). Predicting individual differences in attention, memory, and planning in first graders from experiences at home, child care, and school. *Developmental Psychology, 41*, 99-114. doi: http://dx.doi.org/10.1037/0012-1649.41.1.99

*Noble, K.G., McCandliss, B.D., & Farah, M.J. (2007). Socioeconomic gradients predict individual differences in neurocognitive abilities. *Developmental Science, 10*, 464-480. doi: 10.1111/j.1467-7687.2007.00600.x

Noble, K. G., Norman, M. F., & Farah, M. J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental Science*, 8, 74-87. doi:10.1111/j.1467-7687.2005.00394.x

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics, 8*, 157–159. doi: http://dx.doi.org/10.2307/1164923

*Phillipson, S. (2009). Context of academic achievement: lessons from Hong Kong. *Educational Psychology: An International Journal of Experimental Educational Psychology, 29,* 447-468.

*Pinard, F.A. (2011). *Moderational model investigating child temperament, executive functioning, and contextual predictors of externalizing behaviors in preschoolers.* (Unpublished doctoral dissertation). The University of Southern Mississippi, Hattiesburg, MS

Pollak, S.D., Nelson, C.A., Schlaak, M.F., Roeber, B.J., Wewerka, S.S., Wiik, K.L., Frenn, K.A., Loman, M.M., and Gunnar, M.R. (2010). Neurodevelopmental effects of early deprivation in post-institutionalized children. *Child Development*, 81(1), 224-236.

*Raver, C.C., McCoy, D.C., Lowenstein, A.L., & Pess, R. (2012). Predicting individual differences in low-income children's executive control from early to middle childhood. *Developmental Science*. 1-14.

Reardon, S.F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In R. Murnane & G. Duncan (Eds.), *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children*. New York: Russell Sage Foundation Press.

*Rhoades, B. (2011). Examining the link between preschool social-emotional competence and first grade academic achievement: The role of attention skills. *Early Childhood Research Quarterly, 26,* 182-191. doi: http://dx.doi.org/10.1016/j.ecresq.2010.07.003

*Rhoades, B., Greenberg, M. T., & Domitrovich, C. E. (2009). The contribution of inhibitory control to preschoolers' social-emotional competence. *Journal of Applied Developmental Psychology, 30,* 310-320.

Rogers R. D., Kasai K, Koji M, Fukuda R, Iwanami A, Nakagome K., … & Kato, N. (2004). Executive and prefrontal dysfunction in unipolar depression: a review of neuropsychological and imaging evidence. *Neuroscience Research.* 2004; 50(1), 1-11. doi: http://dx.doi.org/10.1016/j.neures.2004.05.003

Romine, C. B. & Reynolds, C. R. (2005). A model of the development of frontal lobe functioning: findings from a meta-analysis. *Applied Neuropsychology*, 12, 190-201.

Rosenthal, R. & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Reviews of Psychology, 52*, 59-82. doi: 10.1146/annurev.psych.52.1.59

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86,* 638-641.

*Sarsour, K. S. *Social inequalities in early neurodevelopment: Mediation and effect modification in population health perspective.* (Unpublished doctoral dissertation). University of California, Berkeley, Berkeley, CA.

*Sarsour, K., Sheridan, M., Jutte, D., Nuru-Jeter, A., Hinshaw, S., & Boyce, W.T., (2011). Family socioeconomic status and child executive functions: the roles of language, home environment, and single parenthood. *Journal of the International Neuropsychological Society, 17,* 120-132. doi: http://dx.doi.org/10.1017/S1041610211001438

Sheridan, M. A., Sarsour, K., Jutte, D., D'Esposito, M., & Boyce, W. T. (2012). The impact of social disparity on prefrontal function in childhood. *Public Library of Science One*, *7*(4), e35744.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75,* 417-453. doi: 10.3102/00346543075003417

Spearman, C. (1910). Correlation calculated from faulty data. British Journal of Psychology, 3, 271–295.

Sterne, J. A., C., Becker, B. J., & Egger, M. (2005). The funnel plot. In H. R. Rothstein, A. J. Sutton, & M. Borentein (Eds.), *Publication Bias in Meta-Analysis* (75-98). West Sussex, Wiley.

Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine, 22,* 2113-2126.

*Turner, K.A., (2010). *Understanding socioeconomic differences in kindergarteners' school success: the influence of executive function and strategic memory.* (Unpublished doctoral dissertation). North Carolina State University, Raleigh, NC.

*Wiebe, S.A., Espy, K.A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: 1. Latent structure. *Developmental Psychology, 44,* 575-587. doi: http://dx.doi.org/10.1037/0012-1649.44.2.575

Williams, P. G. & Tayer, J. F. (2009). Executive functioning and health: Introduction to the special series. *Annals of Behavioral Medicine, 37,* 101-105. doi: 10.1007/s12160-009-9091-x

Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M. & The Family Life Project Investigators (2010). The measurement of executive function at age 3 years: psychometric properties and criterion validity of a new battery of tasks. *Psychological Assessment*, *22*(2), 306-317. doi: http://dx.doi.org/10.1037/a0018708

Figure Legends

Figure 1. Flow chart illustrating the identification of included studies.

Figure 2. Forest Plot of all studies included in the meta-analysis.

Figure 3. Scatterplot of the relationship between number of EF measures and effect size.

Figure 4. Scatterplot of the relationship between number of SES measures and effect size.

Figure 5. Funnel Plot of Standard Error by Fisher's Z for all studies included in the meta-analysis.

Table 1

For each independent sample, the coding decision about whether meaningful SES variability was reported, the information used to determine this, and the page number of this information is shown

Study	Meaningful SES variability reporte	Information used to determine amount of SES variability	Page
Berry, D., Blair, C., Willoughby, M., Granger, D., & The Family Life Project Key Investigators. (2012). & Blair, C., et al. (2011).	YES	Blair et al. (2011): Table 1	Blair et al. (2011): 1976
		Cameron et al. (2012): "This study examined the contribution of executive function (EF) and multiple aspects of fine motor skills to achievement on 6 standardized assessments in a sample of middle-socioeconomic status kindergarteners."	Cameron et al. (2012): 1233 McClelland et al. (2007): 950
Cameron, C. et al (2012) & McClelland et al. (2007)	Cameron et al., 2012: NO McClelland et al., 2007: YES; Therefore Excluded	McClelland et al. (2007): "Children were recruited from two sites: a predominantly middle to upper-middle-SES urban fringe area with a range of economic and ethnic diversity in Michigan, and a mixed-SES rural site in Oregon."	
De Jong, P. F. (1993)	NO	Not Reported	
		Deng (2008): "Participants (206 families) were recruited at 3 months of child age for the Durham Child Heath and Development (DCHD) Study. Distributed across levels of income and education, these families came from the greater Durham area in North Carolina, and were specifically recruited to represent an approximately equal number of European and African American families. Formal education among the participating mothers varied widely, with 14% 24 having no high school degree, 18% having a high school diploma or G.E.D., 22% with some college or vocational school experience, 29% with a four-year bachelor's degree, and 17% having received education beyond a bachelor degree."	Deng (2008): 23-24 Turner (2010): 10
Deng, M. (2008). & Turner, K.A., (2010).	YES	Turner (2010): "At the 60-month visit, family income-to-needs ratios ranged from 0.05 to 24.33 with an average of 4.41 (SD = 3.39). Mother's years of education for the sample ranged from 5 to 20 years with an average of 15.09 (2.61) years or at least some college education."	

Deprince, A.P., Winzierl, K.M., & Combs, M.D. (2009).	NO	Not Reported	
		Dilworth-Bart et al. (2008): "On average, mothers attained some college level education by the participant child's birth (mean = 14:26; S:D: = 2:50; range = 7221)."; Income-to-need data from 1, 6, 15 and 24 months were averaged for this analysis (mean = 3:29; S:D: = 2:26; range = 0:09–19.29)."	Dilworth-Bart et al. (2008)
		Jacobson (2008): "On average, mothers had 14.44 years of education, with a range of 7 to 21 years. Although this sample generally consisted of well educated mothers, 71 (7.7%) of the mothers did not finish high school. The average family income-to-needs ratio (based on US census definitions for poverty levels) for this sample was 3.45, with a range from .02 to 23.68"	Jacobson (2008): 45
Dilworth-Bart, J.E , Khurshid, A., Lowe Vandell, D., (2007). & Jacobson, L., (2008) & NICHD Early Child Care Research Network (2005).	YES	NICHD Early Child Care Research Network (2005): "Data from the NICHD Study of Early Child Care and Youth Development were used to address the research questions raised above. The NICHD study is a prospective longitudinal study of a large, geographically, ethnically, and economically diverse sample of children born in 1991 and their families."	NICHD Early Child Care Research Network (2005): 101
		"Mean household income was $55,911.08 (SD = 43,125.56); Four (8.2%) mothers completed some high school, 13 (26.5) obtained a high school diploma or equivalent, six (12.2%) obtained a trade or vocational degree, 19 (38.8%) obtained a bachelor's or associate's degree, and seven (14.3%) obtained a graduate or professional degree."	418
Dilworth-Bart, J.E., (2012).	YES		
Doan, S.N. & Evans, G.W. (2011).	YES	Table 2	18
Fernald, L. et al. (2011).	YES	Table 1	837
Hackman, D.A. (2012). Chapter 2.	YES	Table 2 reports parental education (range for primary caregiver: 3-25 years)	114
Henning, A., Spinath, F.M., Aschersleben, G. (2010).	NO	Not Reported	
		"With respect to parent income, 22 (31%) of them were from a low-income level (699 TL and below), 26 (36.6%) of them from a mid-income level (700-1999TL), and 23 (32.4%) of them from a high-income level (2000 TL and above)."	241-242
Ivrendi, A. (2011)	YES		
Kegel, C.A. & Bus, A.G. (2012).	NO	"Participants were 312 kindergartners (60% male) from 15 Dutch schools	184

in Rotterdam, Leiden, and the surrounding areas. Schools were selected for inclusion if they served large numbers of low- SES families and agreed to participate. For 70% of the mothers in our sample, their highest level of education was senior secondary
vocational education (about 13 years of education, excluding prekindergarten)."

Knipe, H., (2009).	NO	Not Reported	
Li-Grining, C. (2005)	NO	Table 2.1	42
		Matte-Gangé & Bernier (2011): "Family income varied from less than $20,000 CDN to more than $100,000 CDN, with an average of $70,000 CDN. Mothers were predominantly Caucasian (86% of sample) and French speaking (79% of sample). They were between 24 and 45 years old (M = 31.2). They had between 10 and 18 years of formal education (M = 15), and 55.8% had a college degree."	Matte-Gangé & Bernier (2011): 614
*Matte-Gangé, C. & Bernier, A. (2011). & Bernier et al. (2012)	YES		
		Mezzacappa et al. (2011): "Participants were 249 children (47% female; 54% Hispanic, 24% African-American, 22% Caucasian) from a	Mezzacappa et al. (2011): 883
Mezzacappa, E., Buckner, J.C., & Earls, F. (2011) & Mezzacappa, E. (2004).	YES	Mezzacappa et al. (2011): "Participants were 249 children (47% female; 54% Hispanic, 24% African-American, 22% Caucasian) from a wide range of SES backgrounds who were followed from infancy in the Project on Human Development in Chicago Neighborhoods (PHDCN)"	
Noble, K.G., McCandliss, B.D., & Farah, M.J. (2007).	YES	"...the mean income-to- needs ratio in our sample of 130 parents who provided this information was 3.36 (SD 3.78); however, whereas the minimum ratio was only 0.23 (less than one standard deviation from the mean), the maximum was 19.5 (over 4 standard deviations from the mean)."	469
Phillipson, S. (2009)	YES	Figure 1	455
		"The sample represented a diverse socioeconomic status background. Scores on the Hollingshead Four Factor Index of Social Status ranged from 14 to 66 (M = 39.15, SD = 15.08)"; Table 2 (Caregiver's Educational Level); The total household yearly income for the current	48-49
Pinard, F. (2011).	YES	sample ranged from "earns no income/dependent on welfare" to "earns	

		over $100,000"	
Raver, C.C., McCoy, D.C., Lowenstein, A.L., & Pess, R. (2012).	NO	"Children resided in families with an average income-to-needs ratio in elementary school of 0.83 (SD = 0.76), indicating that the majority of children in this sample came from families whose annual income and family size placed them below the federal poverty line (which is equal to 1.00)."; Table 1	397-398
Rhoades, B., Greenberg, M., & Domitrovich, C. (2009).	NO	Table 1	313
Rhoades, B., Warren, H. K., Domitrovich, C. E., & Greenberg M. T. (2011).	NO	"The data for the present study come from an economically disadvantaged sample of children (n = 341) in a public preschool program in an urban school district in the Northeastern United States across three years."; Table 1	184
Sarsour, K., Sheridan, M., Jutte, D., Nuru-Jeter, A., Hinshaw, S., & Boyce, W.T., (2011). & Sarsour, K. S. (2007).	YES	Sarsour et al. (2011): "A community sample of 60 families (from a wide spectrum of socioeconomic backgrounds) was recruited from the San Francisco Bay Area…"	Sarsour et al. (2011): 122
Wiebe, S.A., Espy, K.A., & Charak, D. (2008).	YES	"The average maternal education of the sample was 14 years 1 month (SD 2 years 3 months; range: 8 years to 20 years)."	577

Table 2.

Characteristics of the papers (published and unpublished manuscripts) used in the meta-analysis.

Publication	N (total sample)	Country	% Male	Predominant Race	Age range	Mean age	Intended sample SES	Meaningful SES variability reported?	Extent of exclusionary criteria	Type of publication	SES as a focus
Bernier, A., Carlson, S., Deschênes, M., & Matte-Gagné, C. (2012). [1]	62	Canada	38.70	WHITE	<2	3.08	CONV	YES	HIGH	PUB	YES
Berry, D., Blair, C., Willoughby, M., Granger, D., & The Family Life Project Key Investigators. (2012). [2]	1292	USA	NR	MIXED	<2	3	LOW	YES	MIN	PUB	NO
Blair, C., et al. (2011). [2]	1292	USA	NR	MIXED	<2	3	LOW	YES	MIN	PUB	YES
Cameron, C. et al (2012). [3]	213	USA	47.00	MIXED	<2	5.82	CONV	NO	MIN	PUB	NO
De Jong, P. F. (1993)	376	NETHERLANDS	NR	WHITE	<2	9	CONV	NO	NR	PUB	NO
Deng, M. (2008). [4]	206	USA	51.50	MIXED	<2	3	REP	YES	NR	UNPUB	NO
Deprince, A.P., Winzierl, K.M., & Combs, M.D. (2009).	114	USA	42.00	MIXED	2-3.99	10.39	CONV	NO	MIN	PUB	NO
Dilworth-Bart, J.E., (2012).	49	USA	53.06	WHITE	<2	5	CONV	YES	MIN	PUB	YES
Dilworth-Bart, J.E., Khurshid, A., Lowe Vandell, D., (2007). [5]	1273	USA	52.00	WHITE	<2	4.33	REP	YES	HIGH	PUB	YES
Doan, S.N. & Evans, G.W. (2011).	342	USA	NR	WHITE	NR	17.29	LOW	YES	NR	PUB	NO
Fernald, L. et al. (2011).	1332	MADAGASCAR	47.60	NR	2-3.99	4.55	REP	YES	MIN	PUB	YES
Hackman, D.A. (2012). Chapter 2.	316	USA	45.90	MIXED	2-3.99	13.52	REP	YES	HIGH	UNPUB	YES
Henning, A., Spinath, F.M., Aschersleben, G. (2010).	195	GERMANY	48.70	NR	2-3.99	4.92	CONV	NO	HIGH	PUB	NO
Ivrendi, A. (2011)	71	TURKEY	50.70	NR	<2	6.0	CONV	YES	MIN	PUB	YES
Jacobson, L., (2008). [5]	925	USA	48.50	WHITE	<2	8.33	REP	YES	HIGH	UNPUB	NO
Kegel, C.A. & Bus, A.G. (2012).	312	NETHERLANDS	60.00	NR	<2	4.4	LOW	NO	MIN	PUB	NO
Knipe, H., (2009).	132	USA	40.90	WHITE	2-3.99	8.95	CONV	NO	MIN	UNPUB	NO
Li-Grining, C. P. (2005)	439	USA	55.00	MIXED	2-3.99	4.50	LOW	NO	MIN	UNPUB	YES
Matte-Gagné, C. & Bernier, A. (2011). [1]	53	CANADA	35.80	WHITE	<2	3.08	CONV	YES	HIGH	PUB	NO

Publication	N (total sample)	Country	% Male	Predominant Race	Age range	Mean age	Intended sample SES	Meaningful SES variability reported?	Extent of exclusionary criteria	Type of publication	SES as a focus
McClelland, M. M. et al. (2007). [3]	310	USA	51.3	WHITE	<2	4.71	CONV	YES	MIN	PUB	NO
Mezzacappa, E. (2004).[6]	249	USA	52.60	MIXED	2-3.99	5.96	REP	YES	MIN	PUB	YES
Mezzacappa, E., Buckner, J.C., & Earls, F. (2011).[6]	249	USA	52.60	MIXED	2-3.99	6.41	REP	YES	MIN	PUB	NO
NICHD Early Child Care Research Network (2005). [5]	727	USA	50.60	WHITE	<2	6.98	REP	YES	HIGH	PUB	NO
Noble, K.G., McCandliss, B.D., & Farah, M.J. (2007).	168	USA	53.30	MIXED	<2	6.5	REP	YES	MIN	PUB	YES
Phillipson, S. (2009)	215	HONG KONG	47.90	NR	4-6.99	10.70	CONV	YES	MIN	PUB	NO
Pinard, F. (2012).	138	USA	50.70	MIXED	2-3.99	4.03	LOW	YES	MIN	UNPUB	YES
Raver, C.C., McCoy, D.C., Lowenstein, A.L., & Pess, R. (2012).	391	USA	45.50	BLACK	2-3.99	4.2	LOW	NO	MIN	PUB	YES
Rhoades, B. (2011).	341	USA	47.00	BLACK	<2	5.67	LOW	NO	MIN	PUB	NO
Rhoades, B., Greenberg, M., & Domitrovich, C. (2009).	146	USA	46.00	MIXED	<2	4.5	LOW	NO	MIN	PUB	NO
Sarsour, K. S. (2007).[7]	60	USA	31.70	MIXED	4-6.99	9.9	REP		HIGH	UNPUB	YES
Sarsour, K., Sheridan, M., Jutte, D., Nuru-Jeter, A., Hinshaw, S., & Boyce, W.T., (2011) [7]	60	USA	31.70	MIXED	4-6.99	9.9	REP	YES	HIGH	PUB	YES
Turner, K.A., (2010). [4]	138	USA	47.10	MIXED	<2	5	REP	YES	NR	UNPUB	YES
Wiebe, S.A., Espy, K.A., & Charak, D. (2007).	243	USA	44.40	WHITE	2-3.99	3.92	CONV	YES	MIN	PUB	NO

Note. NR = not reported; WHITE = > 60% of the sample identified as 'White;' BLACK = >60% of the sample identified as 'Black;'

MIXED = no single racial or ethnic group made up > 60% of the sample; CONV = convenience sample, REP = representative or

diverse sample, LOW = predominantly low-SES sample, MIN = minimal health- and performance-based exclusionary criteria; HIGH

= high health- and performance-based exclusionary criteria

Table 3.

Effect size information for the samples used in the meta-analysis

Study	Number of ES reported	Pearson's *r* (Average)	*N* (Average)	SES measures	EF measures	SES constructs				EF constructs				
						Edu	Inc	Occ	SES	WM	AS	In	EF	Other
Bernier, A., Carlson, S., Deschênes, M., & Matte-Gagné, C. (2012). & Matte-Gange, C. & Bernier, A. (2011).	1	.38	57.5	SES composite from: Maternal education, Paternal education, Family income	Day/Night, Dimensional Change Card Sort, Bear/Dragon				x					x
Berry, D., Blair, C., Willoughby, M., Granger, D., & The Family Life Project Key Investigators. (2012). & Blair, C., et al. (2011).	4	.35	1121	Income-to-needs, maternal education, caregiver education	EF composite from: Item selection attention shifting, Spatial Conflict inhibitory control, span-like working memory task	x	x						x	
Cameron, C. et al (2012). & McClelland et al. (2007).	4	.11	242.5	Maternal education; Parental education	Heads-Shoulders-Knees-Toes, Head-to-Toes Task	x						x		
De Jong, P. F. (1993)	9	.08	376	Maternal education, paternal education, Paternal occupation	Digit Span, Star Counting, Syllable Counting	x		x		x	x			
Deng, M. (2008). & Turner, K.A., (2010).	9	.16	138	Maternal education, Income-to-needs ratio	Flexible Item Selection Task, Day/Night, backwards Digit Span,	x	x			x	x	x		
Deprince, A.P., Winzierl, K.M., & Combs, M.D. (2009).	1	.20	110	Hollingshead occupational prestige, Parental education, parental occupation	WISC arithmetic, letter-number sequencing, & digit span, symbol search, Gordon Diagnostic System, Stroop task				x				x	

META-ANALYSIS OF SES AND EF

Study	Number of ES reported	Pearson's r (Average)	N (Average)	SES measures	EF measures	SES constructs				EF constructs				
						Edu	Inc	Occ	SES	WM	AS	In	EF	Other
Dilworth-Bart, J.E , Khurshid, A., Lowe Vandell, D., (2007). & Jacobson, L., (2008) & NICHD Early Child Care Research Network (2005).	9	.19	857	Family income, maternal education, income-to-needs, income category	Continuous Performance Test, Day/Night, WJ-R Memory for Sentences, Tower of Hanoi, Delay of gratification	x	x						x	x
Dilworth-Bart, J.E., (2012).	12	.36	49	Maternal education, Household income, SES	Peg-tapping, Fish Flanker, Stanford-Binet verbal & non-verbal working memory	x	x		x	x	x	x	x	
Doan, S.N. & Evans, G.W. (2011).	1	.14	214	Poverty	Working memory (spatial task)		x			x				
Fernald, L. et al. (2011).	4	.18	1064	Maternal education, paternal education	Working Memory, Leiter-Revised Attention Sustained Task	x				x				x
Hackman, D.A. (2012). Chapter 2.	7	.14	314	Parental education	Corsi Block Tapping, Digit Span Backwards, Spatial Working Memory, Object 2-back, Stop Signal Reaction Time, Stroop, Flanker	x				x		x		
Henning, A., Spinath, F.M., Aschersleben, G. (2010).	2	.17	175	Maternal education, paternal education	Dimensional Change Card Sort	x						x		
Ivrendi, A. (2011)	3	.47	70	Maternal education, family income	Head, Toes, Knees & Shoulders (HTKS) Stroop-like task (opposites), Stroop-like task (dogs), Digit span (words), WISC Backward Digit Span	x	x					x		
Kegel, C.A. & Bus, A.G. (2012).	2	.12	283	Maternal education	Backward Digit Span	x				x		x		

Study	Number of ES reported	Pearson's r (Average)	N (Average)	SES measures	EF measures	SES constructs				EF constructs				
						Edu	Inc	Occ	SES	WM	AS	In	EF	Other
Knipe, H., (2009).	3	.06	125	Parental education	Backward Digit Span, Wisconsin Card sort, D-KEFS tower	x				x	x			x
Li-Grining, C. (2005)	4	.04	438	Maternal education (less than HS vs. HS and above); Income-to-needs	Shapes, Turtle/Rabbit	x	x					x		x
Mezzacappa, E., Buckner, J.C., & Earls, F. (2011) & Mezzacappa, E. (2004).	6	.20	233	SES	Flanker				x		x			
Noble, K.G., McCandliss, B.D., & Farah, M.J. (2007).	2	.24	150	SES composite	Spatial working memory task, Delayed nonmatch to sample, Go/no-go task, NEPSY auditory attention and response set				x	x		x		
Phillipson, S. (2009)	1	.21	215	SES	Swanson Cognitive Processing Test				x	x				
Pinard, F. (2011).	1	-.04	102	Hollingshead Four Factor Index of Social Status	Day/night, Grass/snow, NEPSY-II Statue, NEPSY-II Auditory Attention				x				x	
Raver, C.C., McCoy, D.C., Lowenstein, A.L., & Pess, R. (2012).	4	.02	328	Mother < HS education, Income-to-needs	Balance Beam, Pencil Tap	x	x					x	x	x
Rhoades, B., Greenberg, M., & Domitrovich, C. (2009).	3	-.02	131	Primary caregiver education	Leiter-Revised Attention Sustained subtest, Day/Night, Peg Tapping	x						x		x
Rhoades, B., Warren, H. K., Domitrovich, C. E., & Greenberg M. T. (2011).	2	.11	288	Maternal education, Family income	Leiter-Revised Attention Sustained Task	x	x							x

Study	Number of ES reported	Pearson's *r* (Average)	*N* (Average)	SES measures	EF measures	SES constructs				EF constructs				
						Edu	Inc	Occ	SES	WM	AS	In	EF	Other
Sarsour, K., Sheridan, M., Jutte, D., Nuru-Jeter, A., Hinshaw, S., & Boyce, W.T., (2011). & Sarsour, K. S. (2007).	6	.37	60	Income-to-needs ratio, Hollingshead Index of Occupational Status, Family wealth, Maternal education	WISC Digit Span, Trail Making Test, Stroop Test Delayed Attention, DAS Digit Span, Six Boxes, Delayed Response, NEPSY Statue, Whisper, Child Continuous Performance Task, shape school, NEPSY Visual Attention, Tower of Hanoi				x	x	x	x		
Wiebe, S.A., Espy, K.A., Charak, D. (2008).	10	.14	198	Maternal education		x				x		x		

Note. Correlations and sample sizes were averaged across all effect sizes reported in each study.

Edu = education; Inc = income; Occ = Occupation; SES = composite SES measure; WM = working memory; AS = attention shifting; In = inhibition; EF = composite EF; Other = other EF (e.g., planning, sustained attention)

Table 4.

Results of moderation tests for categorical sample characteristics using mixed effects models and fixed effect models.

Moderator	N	r	95% CI	Q(df)	p	r	95% CI	Q_b(df)	p
			Mixed Effects Model				Fixed Effects Model		
Intended sample SES				2.24(2)	.33			.63(2)	.73
Low SES	9	.11	.01- .21			.18	.15-.21		
Representative/Diverse SES	6	.19	.15- .24			.19	.15-.24		
Convenience sample	10	.19	.12- .26			.16	.12-.21		
Amount of SES variability				14.79 (1)**	< .001			35.58(1)	< .001
Meaningful variability reported	15	.22	.16- .28			.23	.20-.25		
Meaningful variability not reported	9	.08	.04- .12			.08	.04-.12		
Extent of exclusionary criteria				1.01 (1)	.32			.37(1)	.54
Minimal	18	.15	.08-.22			.18	.15-.20		
High	5	.20	.13-.26			.19	.14-.24		
Racial composition				3.05 (2)	.22			13.82(2)	.001
>60% White	7	.16	.09- .22			.16	.11-.20		
>60% Black	2	.06	-.03-.16			.06	-.02-.14		
Mixed, none > 60%	10	.17	.06- .27			.22	.18-.25		
Age Range				5.25 (2)	.07			15.92(2)	< .001
< 2 years	13	.19	.12- .27			.21	.18-.24		
2 – 3.99 years	10	.12	.07- .17			.12	.09-.16		
4 - 6.99 years	2	.26	.11-.39			.25	.13-.36		

Note. * $p < .05$. ** $p < .01$.

Table 6.

Results of moderation tests for publication characteristics using mixed effects models and fixed effect models.

Moderator	N	Mixed Effects Model				Fixed Effects Model			
		r	95% CI	Q(df)	p	r	95% CI	Q_b(df)	p
Type of publication				5.05(1)	.03			12.60(1)**	<.001
Published	18	.18	.12-.24			.19	.17-.22		
Unpublished	5	.08	.01-.14			.08	.02-.14		
SES as a primary focus				.75(1)	.39			.26(1)	.61
Yes	9	.17	.08-.25			.14	.10-.18		
No	12	.13	.09-.17			.13	.09-.17		

Note. * $p < .05$. ** $p < .01$.
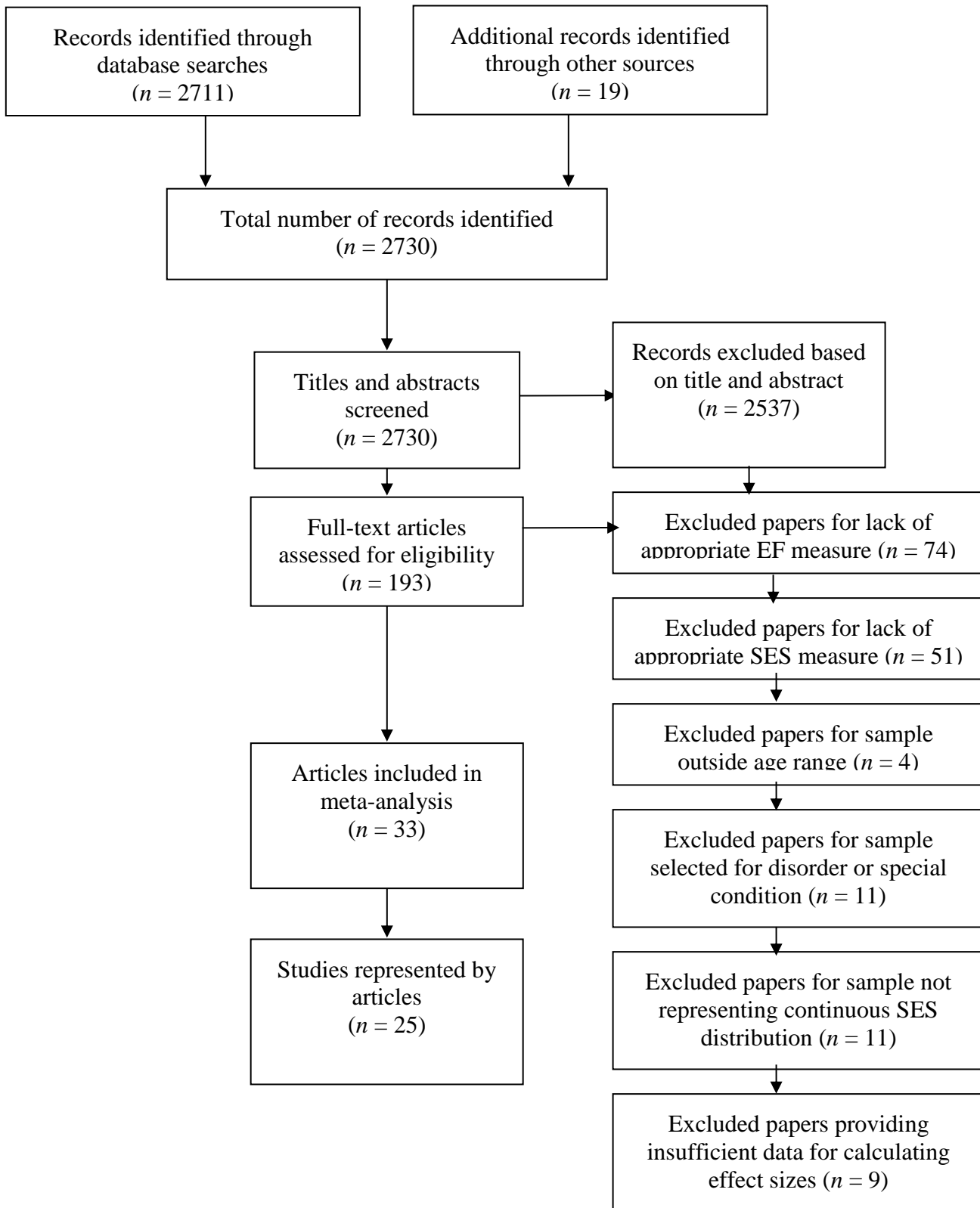
Figure 1. Flow chart illustrating the identification of included studies

Forest Plot of All Studies

STUDYID	Statistics for each study					Correlation and 95% CI
	Correlation	Lower limit	Upper limit	Z-Value	p-Value	
Blair et al. (2011) & Berry et al. (2012)	0.355	0.302	0.406	12.263	0.000	
Cameron (2012) & McClelland et al. (2007)	0.109	-0.023	0.237	1.626	0.104	
De Jong (1993)	0.078	-0.023	0.178	1.507	0.132	
Deprince et al. (2009)	0.200	0.013	0.373	2.097	0.036	
Dilworth-Bart (2012)	0.361	0.089	0.583	2.563	0.010	
Doan & Evans (2011)	0.135	0.001	0.264	1.973	0.048	
Fernald (2011)	0.175	0.116	0.233	5.708	0.000	
Hackman (2012)	0.145	0.035	0.252	2.580	0.010	
Henning et al. (2010)	0.170	0.022	0.311	2.246	0.025	
Ivrendi (2011)	0.472	0.267	0.637	4.201	0.000	
Jacobson (2008), Dilworth-Bart et al. (2007) & NICHD Research Network (2005)	0.190	0.124	0.255	5.578	0.000	
Kegel & Bus (2012)	0.115	-0.001	0.229	1.937	0.053	
Knipe (2009)	0.065	-0.112	0.238	0.719	0.472	
Li-Grining (2005)	0.037	-0.057	0.130	0.763	0.445	
Matte-Gagne & Bernier (2011) & Bernier et al. (2012)	0.376	0.128	0.579	2.905	0.004	
Mezzacappa et al. (2011) & Mezzacappa (2004)	0.195	0.069	0.316	3.010	0.003	
Noble, McCandliss, & Farah (2007)	0.237	0.079	0.382	2.923	0.003	
Phillipson (2009)	0.214	0.083	0.338	3.165	0.002	
Pinard (2011)	-0.043	-0.236	0.153	-0.428	0.669	
Raver et al. (2012)	0.018	-0.091	0.126	0.316	0.752	
Rhoades et al. (2011)	0.114	-0.002	0.226	1.923	0.055	
Rhoades et al. (2009)	-0.023	-0.195	0.150	-0.255	0.799	
Sarsour et al. (2011) & Sarsour (2007)	0.370	0.127	0.570	2.925	0.003	
Turner (2010) & Deng (2008)	0.158	-0.009	0.317	1.857	0.063	
Wiebe, Espy & Charak (2008)	0.141	0.000	0.277	1.961	0.050	
	0.178	0.155	0.200	15.202	0.000	

-1.00   -0.50   0.00   0.50   1.00

Negative Correlation   Positive Correlation

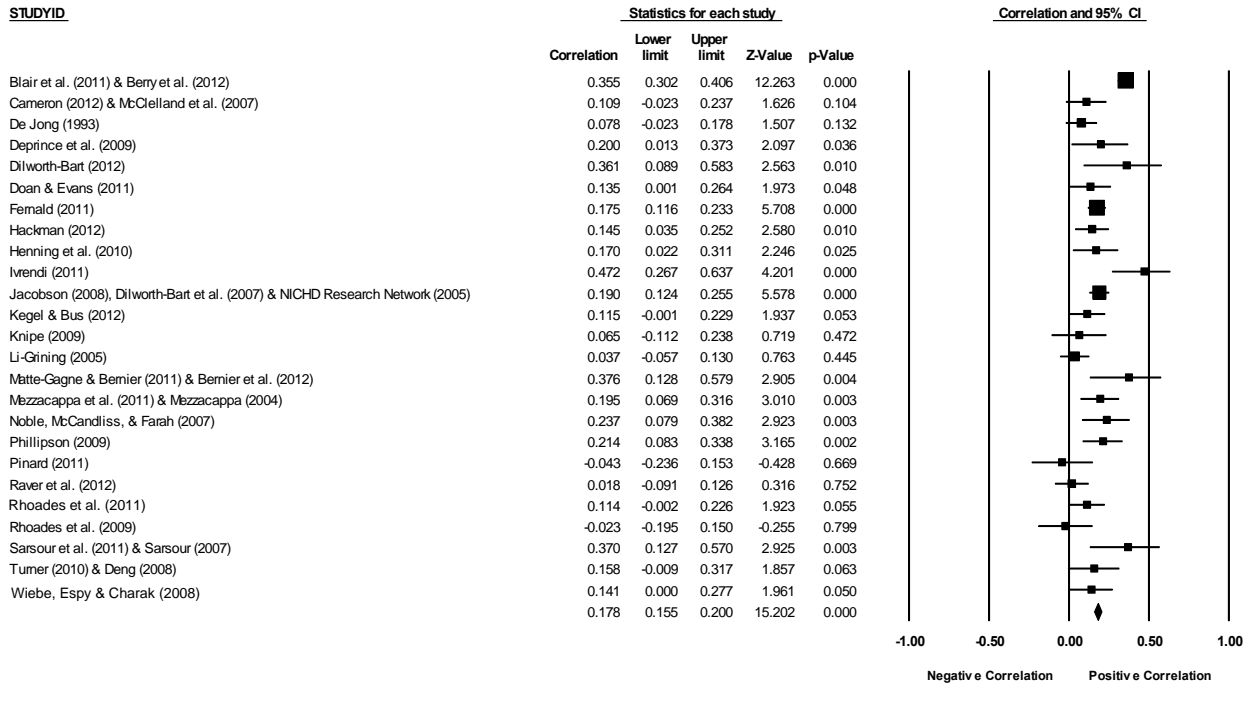*Figure 2.* Forest Plot of all studies included in the meta-analysis.

Figure 3. Scatterplot of the relationship between number of EF measures and effect size.
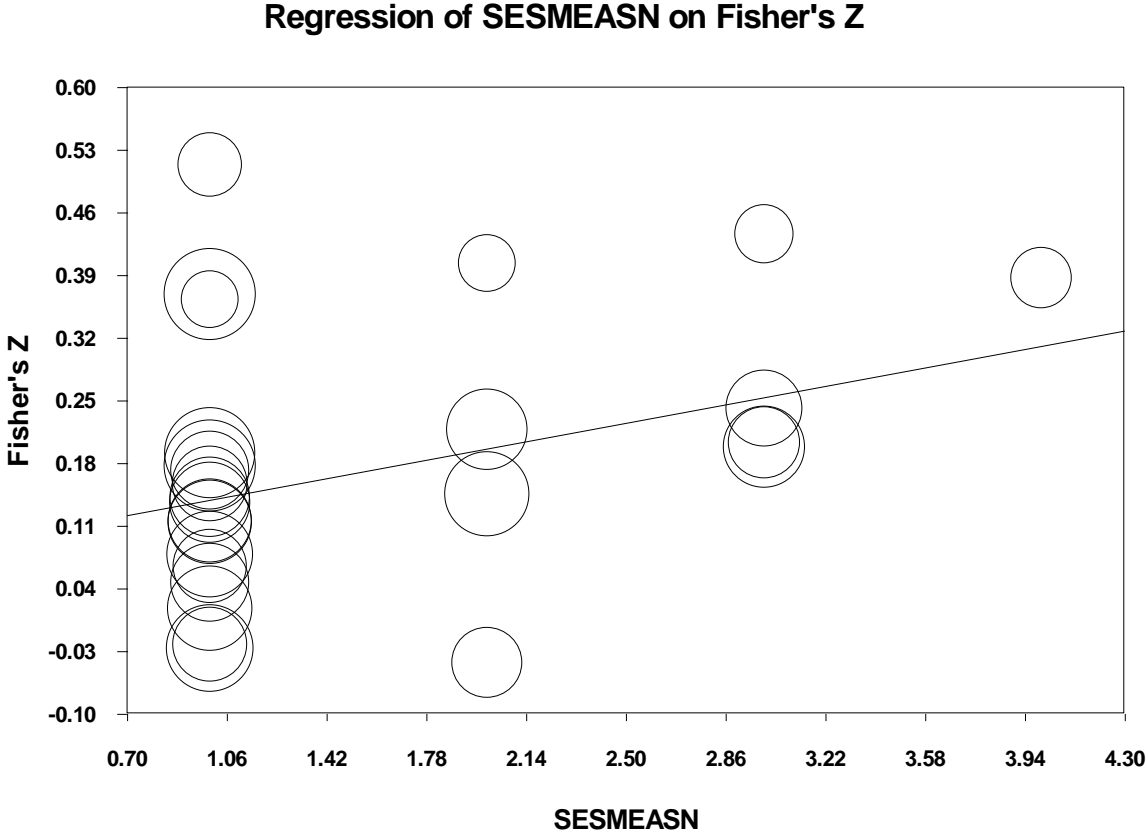
Regression of SESMEASN on Fisher's Z

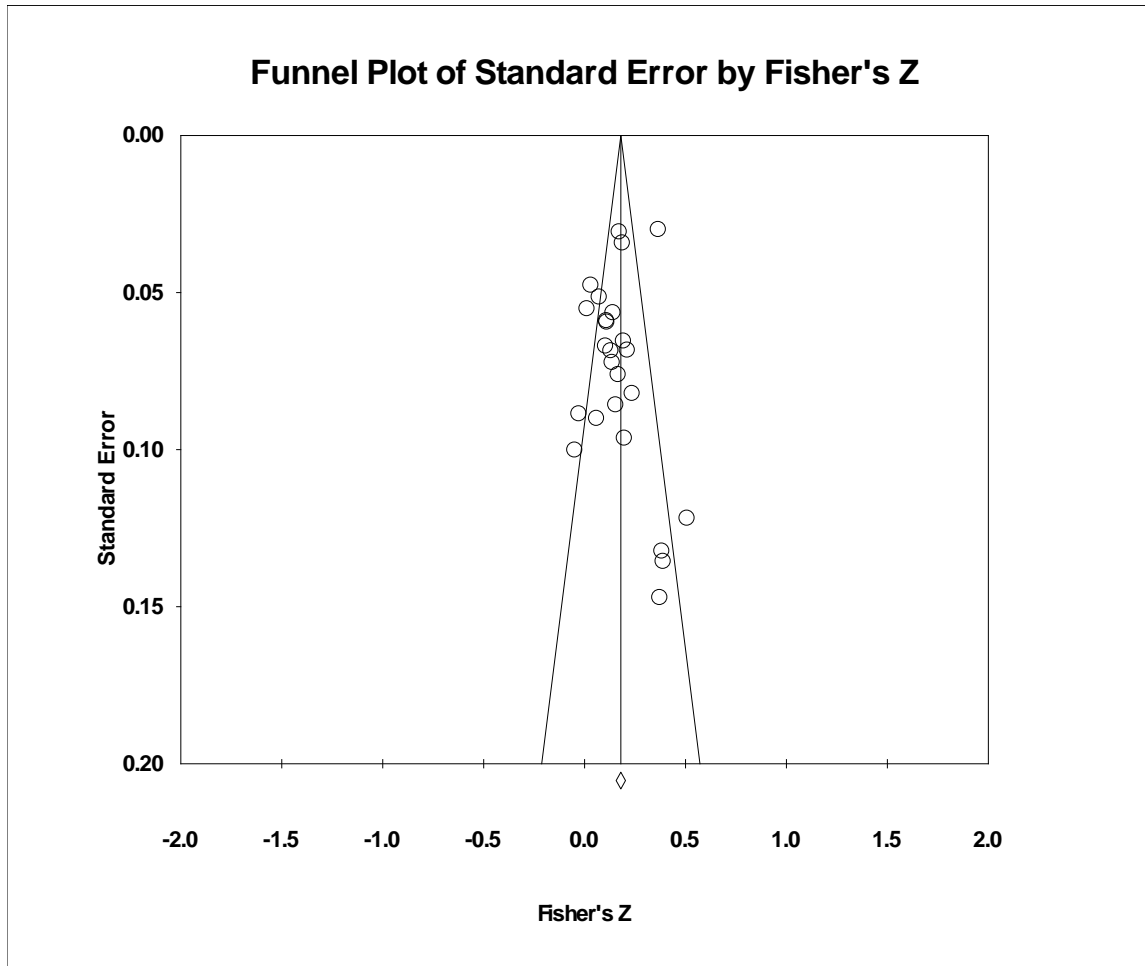*Figure 4.* Scatterplot of the relationship between number of SES measures and effect size.

Figure 5. Funnel Plot of Standard Error by Fisher's Z for all studies included in the meta-analysis.