Open access • Journal Article • DOI:10.1007/S10964-019-00982-8

# The Effects of Youth Mentoring Programs: A Meta-analysis of Outcome Studies.



Elizabeth B. Raposa, Jean E. Rhodes, Geert Jan J. M. Stams, Noel A. Card ...+6 more authors

**Institutions:** College of William & Mary, University of Massachusetts Boston, University of Amsterdam, University of Connecticut ...+3 more institutions

**Published on:** 19 Jan 2019 - Journal of Youth and Adolescence (J Youth Adolesc)

**Topics:** Youth mentoring and Positive Youth Development

# The Effects of Youth Mentoring Programs: A Meta-analysis of Outcome Studies

Raposa, E.B.; Rhodes, J.; Stams, G.J.J.M.; Card, N.; Schwartz, S.; Yoviene Sykes, L.A.; Kanchewa, S.; Kupersmidt, J.; Hussain, S.

**DOI**
10.1007/s10964-019-00982-8

**Publication date**
2019

**Document Version**
Final published version

**Published in**
Journal of Youth and Adolescence

Journal of Youth and Adolescence (2019) 48:423–443
https://doi.org/10.1007/s10964-019-00982-8

SYSTEMATIC LITERATURE REVIEW

# The Effects of Youth Mentoring Programs: A Meta-analysis of Outcome Studies

Elizabeth B. Raposa[1] · Jean Rhodes[2] · Geert Jan J. M. Stams[3] · Noel Card[4] · Samantha Burton[2] · Sarah Schwartz[5] · Laura A. Yoviene Sykes[2] · Stella Kanchewa[2] · Janis Kupersmidt[6] · Saida Hussain[7]

Received: 8 October 2018 / Accepted: 5 January 2019 / Published online: 19 January 2019


## Abstract

Mentoring programs, which pair youth with caring, non-parental adults with the goal of promoting positive youth development, are an increasingly popular strategy for early intervention with at-risk youth. However, important questions remain about the extent to which these interventions improve youth outcomes. The present study involved a comprehensive meta-analysis of all outcome studies of intergenerational, one-on-one youth mentoring programs written in the English language between 1975 and 2017, using rigorous inclusion criteria designed to align with developmental theories of youth mentoring. Analysis of 70 mentoring outcome studies, with a sample size of 25,286 youth (average age of 12 years old), yielded a statistically significant effect of mentoring programs across all youth outcomes. The observed effect size fell within the medium/moderate range according to empirical guidelines derived from universal prevention programs for youth, and was consistent with past meta-analyses of youth mentoring. Moderation analyses indicated that programs serving a larger proportion of male youth, deploying a greater percentage of male mentors or mentors with a helping profession background, and requiring shorter meetings yielded larger effect sizes, as did evaluations that relied on questionnaires and youth self-report. Taken together, these findings provide some support for the efficacy of mentoring interventions, while also emphasizing the need to remain realistic about the modest impact of these programs as currently implemented, and highlighting opportunities for improving the quality and rigor of mentoring practices.

**Keywords** Meta-analysis · Youth mentoring · Relational theory

## Introduction

Youth mentoring programs show great promise as a low-cost intervention for youth at risk for developing a range of psychological, social, and behavioral problems. Recent research has highlighted the positive impact of one-on-one mentoring relationships for children and adolescents showing externalizing behaviors such as aggression (Jolliffe and Farrington 2007), substance use (Rhodes et al. 2005), and other delinquent behaviors (Tolan et al. 2008). In addition, one recent study assessed the influence of mentoring relationships on a wide range of youth outcomes, and showed particularly potent effects for mentoring on youth depressive symptoms (Herrera et al. 2013). As a result, youth mentoring programs have grown in popularity as a strategy for intervening with youth at-risk for diverse problems (Blakeslee and Keller 2012), and an estimated 2.5 million U.S. children and adolescents are paired with caring adults through mentoring programs each year (Raposa et al. 2017). Important questions remain, however, about the

**Supplementary information** The online version of this article (https://doi.org/10.1007/s10964-019-00982-8) contains supplementary material, which is available to authorized users.

✉ Elizabeth B. Raposa
ebraposa@wm.edu

[1] Department of Psychological Sciences, College of William and Mary, 540 Landrum Dr., Williamsburg, VA 23188, USA

[2] Department of Psychology, University of Massachusetts, Boston, MA, USA

[3] Department of Child Development and Education, University of Amsterdam, Amsterdam, Netherlands

[4] Department of Human Development and Family Studies, University of Connecticut, Storrs, CT, USA

[5] Department of Psychology, Suffolk University, Boston, MA, USA

[6] Innovation Research & Training, Durham, NC, USA

[7] Department of Psychology, University of Virginia, Charlottesville, VA, USA

extent to which mentoring interventions influence youth outcomes, and the conditions under which they are most effective (e.g., Aos et al. 2004). Many of the existing meta-analyses designed to summarize the effectiveness of mentoring programs have focused on subsets of youth or particular program approaches (e.g., Tolan et al. 2008). Moreover, the most recent meta-analysis of mentoring outcomes included studies that were conducted between 1999 and 2010 (DuBois et al. 2011), and therefore did not capture the recent proliferation of mentoring studies designed to influence an increasingly diverse set of youth outcomes (Blakeslee and Keller 2012).

The current study aims to address these gaps in the existing literature by conducting a comprehensive meta-analysis of all mentoring outcome studies written in the English language to-date, with a focus on intergenerational, one-on-one mentoring programs that are consistent with a developmental conceptual model of youth mentoring. These meta-analytic findings have relevance to researchers and practitioners invested in preventing mental health problems and promoting positive youth development.

## A Developmental Conceptual Model of Youth Mentoring

Although there is considerable diversity in the structure and purpose of adult-youth mentoring interventions, most programs are grounded in a substantial literature showing the importance of supportive intergenerational relationships for promoting positive youth development and preventing a host of psychosocial problems, such as depression and delinquent behavior (DuBois and Karcher 2013). Present analyses draw from a developmental model of youth mentoring relationships (Rhodes et al. 2002; Rhodes 2005) as a guiding conceptual framework. This developmental model posits an interconnected set of three processes (i.e., social-emotional, cognitive, and identity formation processes) through which the establishment of close, caring relationships with non-parental adults are expected to promote positive developmental trajectories.

First, by modeling prosocial skills and providing a consistent and safe relational context, mentors are thought to enhance youth's perceptions of social support and to facilitate more positive connections with others. In particular, relationships with caring adults may become a context for helping youth interpret and manage interpersonal difficulties, improving peer and adult relatedness, and increasing youth's receptivity to adult values, advice, and perspectives (Ruzek et al. 2016). The basis for expecting that positive mentoring relationships can modify youths' perceptions of other relationships is derived largely from attachment theory, which posits that children construct cognitive representations of relationships through their early experiences with primary caregivers, which in turn influence interpersonal behavior (Bowlby 1988). Although these experience-based and generalized expectations around social interactions, or working models of attachment, are relatively stable over time, they remain flexible to modification in response to changing life circumstances, such as engagement in supportive relationships (Belsky and Cassidy 1994). Empirical research on mentoring has provided consistent evidence for such processes, indicating that high-quality mentoring relationships are associated with improvements in social and emotional functioning, including perceptions of relationships with parents, peers, and teachers (Cavell et al. 2013; Kanchewa et al. 2016; Karcher et al. 2002).

Second, youth engagement in shared activities and meaningful conversations with more sophisticated thinkers is thought to scaffold and advance cognitive skills (Rogoff 1978). Key cognitive processes, such as information processing and self-regulation, strengthen during adolescence, particularly in the context of supportive interactions with caring adults (Parra et al. 2002). Research on the role of social support in fostering cognitive development underscores the social nature of learning and, specifically, the potential contributions of adults in mentoring roles. Feelings of closeness with teachers, for example, have been associated with greater cognitive engagement and executive functioning, as well as more positive academic adjustment for children and adolescents (Spilt et al. 2012).

Finally, mentors are thought to promote identity development by serving as concrete models of success, demonstrating qualities that youth might wish to emulate and exposing youth to new contexts and resources for interest exploration (Sánchez et al. 2006). Markus and Nurius (1986) have referred to *possible selves*: individuals' ideas of what they might become, what they would like to become, and what they fear becoming. Such possibilities, which often emerge as youth observe and compare the adults they know, can inform decisions and desired behaviors. Along these lines, mentors can open doors to activities, resources, and educational or occupational opportunities that youth can draw on to construct their sense of identity (Darling et al. 2002). Indeed, findings regarding mentors' protective influence on risk behaviors, and related improvements in physical health and well-being, are suggestive of a more positive future orientation in youth (Herrera et al. 2013).

It is important to note that, according to this developmental model of youth mentoring, improvements across these three domains (i.e., social-emotional, cognitive, and identity formation) are interconnected, and the long-term impact of these improvements on more distal youth outcomes is often non-specific. For example, the use of a mentor as a role model, and the ability to entertain multiple possible selves in the service of identity development, may

be fostered by the ability of the youth to enter into a secure, trusting relationship, as well as a growing capacity to understand the world from the perspective of others. Likewise, growth in certain cognitive abilities, such as information processing and self-regulation, can enhance the capacity of youth to regulate complicated emotions. As a result, mentoring programs often assess, and see improvements in, a wide variety of youth outcomes, even if the proposed intervention mechanism targets only one of these developmental processes.

## Effectiveness of Youth Mentoring

Several meta-analytic studies have advanced the field's understanding of the effect of youth mentoring on various outcomes. Some meta-analyses have focused on specific subsets of youth or particular program models. For example, meta-analyses with youth at risk for delinquent or aggressive behavior have found impacts of mentoring on juvenile reoffending (Cohen's $d = .21$; Jolliffe and Farrington 2007) and delinquency (Cohen's $d = .23$; Tolan et al. 2008). Another meta-analysis of three large-scale, school-based mentoring evaluations showed positive effects of mentoring on a range of school-related outcomes (Cohen's $d$ ranging from .07 to .18; Wheeler et al. 2010). In addition, when mentoring programs were evaluated across youth, academic, and workplace settings, the effect of mentoring was, again, statistically significant, with youth mentoring programs showing somewhat smaller effects on most outcomes (sample-size weighted corrected correlations ($r_c$) ranging from .03 to .14) than mentoring programs for adults implemented in workplace or higher education settings ($r_c$ ranging from .03 to .36; Eby et al. 2008).

It should be noted that, according to traditional conventions for the interpretation of the magnitude of effect sizes (Cohen 1988), a standardized mean difference of 0.20 or lower is considered small (while 0.50 is considered medium/moderate and 0.80 is considered large), which would suggest a fairly limited impact of these youth mentoring interventions. However, it is more informative to interpret effect sizes using guidelines derived empirically from this particular intervention area. One recent review of meta-analyses within the field of universal youth prevention programs showed that the median average effect of programs tended to fall within the range of 0.07 to 0.20 standard deviations, with different effects depending on the outcome assessed (Tanner-Smith et al. 2018). These findings suggest that the average observed effect for youth mentoring in these meta-analyses tends to be medium/moderate and fairly consistent with other programs designed to improve a range of youth outcomes, including externalizing and internalizing behaviors, social competence, drug use, and academic achievement.

DuBois and colleagues have published the most comprehensive meta-analyses of youth mentoring programs to date, both of which showed similarly sized effects of mentoring across outcomes (Hedge's $g$ ranging from .18 to .21; DuBois et al. 2002, 2011). Youth who received mentoring, on average, showed functioning on academic, psychosocial, and career outcomes that was about one fifth of a standard deviation higher than non-mentored youth. The most recent of these meta-analyses included studies that were conducted between 1999 and 2010 (DuBois et al. 2011). Although this meta-analysis provided important direction to the field, a recent proliferation of mentoring studies designed to influence an increasingly diverse set of youth outcomes (Blakeslee and Keller 2012) warrants an updated, comprehensive meta-analysis. The current meta-analysis examined all relevant studies of intergenerational, one-on-one youth mentoring programs, which included studies conducted from 1975 through 2017.

In line with the theoretical framework described above, the current study aimed to improve on past meta-analytic findings by using selection criteria that adhered more strictly to a relationship-focused model of mentoring. To this end, analyses included only studies that evaluated a program aimed at improving youth outcomes through a one-on-one, intergenerational mentoring relationship. The present sample therefore consisted of studies that examined a relationship between a younger mentee and an older, non-parental mentor, rather than programs that involved only group mentoring or peer-to-peer mentoring (although some programs did include high school or college students mentoring elementary school students). Likewise, programs that were structured around mentors serving largely instructional roles or administering specific curricula were excluded. Finally, studies in which mentoring was not a primary, or even secondary component, of the intervention were also excluded. These guidelines ensured that analyses were examining mentoring programs designed to improve youth outcomes through a caring, supportive relationship with a non-parental adult.

These more stringent selection criteria led to a different sample of youth mentoring studies than included in previous meta-analyses. As an example, one study that was included in several past meta-analyses (Barnet et al. 2007) was excluded from the present meta-analysis because it evaluated a curriculum-based home visitation program through which intensively-trained home visitors, with caseloads of up to 15 youth, provided structured skills training to pregnant teenagers. Similarly, the present analyses excluded other previously-included studies that evaluated the effectiveness of a structured "character development" curriculum (Everhart 2000) or academic tutoring as the primary intervention (e.g., Burns et al. 2003; Morrow-Howell et al. 2009), with the mentoring component

largely considered secondary and not consistently delivered to all participating youth. Finally, a number of other evaluations were excluded because they focused on group as opposed to one-to-one mentoring (e.g., Cummings 2010; Gatto 2006; Hanlon et al. 2009; Jent and Niec 2009).

## Examining the Influence of Mentoring Across Youth Outcomes

The current meta-analysis also takes a novel approach to answering the important question of which specific youth outcomes are most strongly influenced by mentoring. Youth mentoring has been deployed to address a wide range of emotional, health, behavioral, academic, and vocational outcomes. Past meta-analyses have therefore attempted to evaluate the effects of mentoring on certain subsets of outcomes and have found varying effect sizes depending on the outcome domain. For example, DuBois and colleagues (2002) found that the effect size for mentoring was only .10, on average, for emotional/psychological outcomes, but doubled in size to .22 for career/employment outcomes. Similarly, a more recent meta-analysis (DuBois et al. 2011) found an effect size close to zero for physical health outcomes, versus approximately .20 for conduct problems.

Although these findings suggest that youth mentoring can differentially affect youth outcomes, the specific areas of impact of mentoring on youth are difficult to discern due to the fact that constructs have been inconsistently defined and assessed across studies. As a result, outcomes across categories are sometimes more conceptually similar than those within a particular category. In one study, for example, prosocial attitudes were grouped with achievement motivation in an "attitudinal/motivational" category (DuBois et al. 2011), rather than with social skills and peer relationships in the "social/interpersonal" category. At the same time, externalizing behaviors and internalizing symptoms are often placed in separate categories (DuBois et al. 2002, 2011), despite the close association and frequent comorbidity of these types of psychopathology in the clinical and developmental psychopathology literatures (Bornstein et al. 2010). As mentoring programs increasingly seek to target specific clinical outcomes, there has been a call for greater precision and conceptual clarity in youth outcomes (Arnold and Cater 2016).

To address this issue, the present study utilized a two-tier system to examine the differential impact of mentoring on youth outcomes. In keeping with the typical targets of mentoring interventions, youth outcomes were grouped into five broad categories, including school functioning, social relationships, health, cognition, and psychological symptoms. In addition, sub-categories were created within each broad category. For example, school functioning outcomes were also coded as relevant to one of three sub-categories: school engagement, academic achievement, or extracurricular activities (see Methods section for more details). These categories were derived from an expert review of the developmental psychopathology and positive youth development literatures, and they allowed for assessment of the effectiveness of mentoring on constructs that are aligned with more recent research on the etiology and prevention of clinical issues, as well as the promotion of well-being, in youth.

In addition to re-coding outcomes using a two-tiered system, the current study also used three-level meta-analysis, which accounts for the statistical dependency among effect sizes within studies, and therefore allows for the inclusion of more than one effect size per study. This approach increases statistical power, accounts for both within- and between-study variability, and facilitates analyses of moderators that might explain either within- or between-study variance (Van den Noortgate et al. 2014). The present three-level meta-analysis was designed to account for the nesting of three types of outcome data (i.e., narrow outcome domains within broad outcome domains within overall study effect sizes), while also allowing for estimates of multiple between-study (e.g., mentoring program characteristics, publication type) and within-study (e.g., questionnaire versus interview assessment) moderators.

## Moderators of Mentoring Effectiveness

The current meta-analysis examined a wide range of youth, mentor, and program characteristics that were considered potential moderators of program effects, given increasing evidence that certain individual and program factors might significantly influence the impact of mentoring. As mentoring programs continue to multiply and use a wide array of program practices, it is crucial to identify which of these practices are most helpful to youth, as well as whether any might actually be harmful.

### Youth characteristics

A number of youth demographic characteristics have been shown to moderate the effects of mentoring. There is some evidence that youth mentoring may be more effective with mid- to late-elementary school-aged children, while mentoring relationships are less close and enduring with adolescent mentees (Kupersmidt et al. 2017a). Youth gender may also influence the impact of mentoring relationships. Male and female mentees tend to be referred to mentoring programs for different reasons, with male referrals more commonly stemming from the need for a male role model, while female mentees are more often referred because of relational challenges with their primary caregivers (Rhodes

et al. 2008). Moreover, there is some indication that, relative to males, females are more prone to co-rumination in their dyadic relationships, a process that could attenuate positive effects (Splendelow et al. 2017). Consistent with these ideas, DuBois and colleagues' (2011) meta-analysis demonstrated that programs serving a greater proportion of male mentees had stronger effects. More recently, a substantial body of research has also investigated the role of youth risk factors, with two previous meta-analyses suggesting that youth with higher levels of baseline risk, particularly environmental risk (e.g., poverty, neighborhood violence) may benefit more from mentoring (DuBois et al. 2011; DuBois et al. 2002). Similarly, a study specifically examining the role of youth socioeconomic status (SES) found that youth from lower SES backgrounds benefitted more from mentoring than more affluent peers (Thompson. et al. 2013). However, it should be noted that another large-scale evaluation showed few differences in impacts between students with greater or lower levels of baseline risk (Herrera et al. 2013).

### Mentor characteristics

Several mentor characteristics have also been linked to the effectiveness of youth mentoring. Research indicates that mentor age may affect program outcomes. Specifically, student mentors who volunteer through high school and college programs tend to be less effective than older volunteer mentors (Herrera et al. 2008; Grossman et al. 2012). In contrast, mentors who have more experience in helping roles or professions (e.g., counselor, social worker, therapist) have been found to be more effective than those with non-helping backgrounds, both in formal mentoring relationships (DuBois et al. 2002) and in naturally-occurring or informal mentoring relationships (Van Dam et al. 2017). Studies of mentor demographic variables, such as race and gender, have yielded less consistent results, with most studies showing no effect of these variables on mentoring relationship outcomes (DuBois et al. 2002, 2011).

### Program characteristics

There is substantial diversity in program practices that are included under the umbrella of youth mentoring, which may have implications for the benefits that youth derive from the intervention. Some programs provide mentor incentives, either in the form of payment or course credit, rather than relying on pure volunteerism. Such practices are based on the assumption that increased fidelity will offset incentive costs, although an earlier meta-analysis failed to find significant differences in effects (DuBois et al. 2011). Other variations in program practices relate to expectations for the mentor and youth, including the expected length of the relationship and recommended activities during mentor-youth meetings. Some studies have linked relationship duration to mentee outcomes, showing the greatest benefits from relationships lasting at least 12 months (Grossman and Rhodes 2002). In contrast, other studies suggest that meeting the expected time commitment for the relationship is more important than the actual length of the relationship (Grossman et al. 2012), while recent meta-analyses have failed to detect differences in program effects based on match length (DuBois et al. 2002, 2011). Importantly, there is also considerable variation in the focus of the mentoring relationship, and thus, the intervention that a particular youth receives. For example, some programs focus on academic or vocational development (Woods and Preciado 2016), while others are more general in their focus. Although all of these types of programs fall under the category of mentoring, the interventions they are delivering can vary significantly in ways that influence the benefits that are derived; yet, little research has systematically compared the effects of these different approaches to youth mentoring.

## Methodological Predictors of Mentoring Effect Sizes

Although typically unexamined in the mentoring literature, an important factor that has been consistently shown to predict effect sizes in meta-analyses from other fields involves the methodological approach of the study. Specifically, research shows that studies employing random assignment yield smaller effect sizes than those employing less rigorous quasi-experimental designs (Cheung and Slavin 2015). Additionally, published studies tend to report greater effect sizes than unpublished reports due to biases in publishing significant results (Cheung and Slavin 2015). One recent study showed that strong, statistically significant results are 40 percent more likely to be published than null results (Franco et al. 2014). It is therefore crucial to account for potential publication biases when conducting a thorough meta-analysis. Such biases can be explored by examining differences in effect sizes between different types of research reports as a function of publication status (e.g., dissertations, research reports, and journal articles), but also by comparing the observed distribution of effect sizes against a theoretical distribution of effect sizes showing no publication bias through funnel plot analysis (Egger et al. 1997).

## The Current Study

To address gaps in the existing literature, the current meta-analysis examined the impact of youth mentoring using all relevant outcome studies of intergenerational, one-on-one youth mentoring programs written in English between 1975

and 2017. Stringent inclusion criteria, in line with a developmental framework for conceptualizing the impact of youth mentoring, ensured that analyses were examining mentoring programs designed to improve youth outcomes through a caring, supportive relationship with a non-parental adult. Using a multilevel meta-analytic approach, the analyses (1) estimated the overall effect size of youth mentoring programs, as well as within- and between-study variability in effect sizes; (2) tested whether the effects of youth mentoring were different across diverse outcome categories (e.g., school-related versus psychological outcomes); (3) examined whether the size of program effects were moderated by key youth characteristics, mentor characteristics, program characteristics, and research design issues; and (4) tested the role of publication bias in the calculated overall effect size.

## Method

### Study Selection

A comprehensive search of the literature published prior to June 2017 was conducted to identify evaluations of mentoring programs. Both computer-based and manual search methods were used to locate studies for the current analysis. The computerized databases utilized were PsycINFO, ERIC, and Proquest. A comprehensive search of each computerized database included the following terms and combinations of terms: *Youth mentoring, Mentor + program, Mentor + evaluation, Mentor + intervention, Mentor + outcomes, Mentor + effects, Mentor + comparison, Big Brother, Big Sister, Protégé + mentor, Apprentice + mentor, Nonparental adult + mentor, Mentor mentee relationship, Mentor + randomized control trial, Mentor + RCT, and Mentor + experimental*. These searches yielded peer-reviewed articles, unpublished dissertations, and technical reports. Prior meta-analyses and qualitative reviews were manually searched to identify additional articles. In addition, websites of several national formal mentoring programs (e.g., Big Brothers Big Sisters of America), organizations that routinely evaluate or fund research in the area of mentoring, and the Office of Juvenile Justice and Delinquency Prevention were used to search for articles and reports with more limited circulation. Finally, additional recommendations on relevant research studies were solicited from experts in the field who have been involved in prior youth mentoring meta-analyses, or who have extensive experience conducting mentoring research.

The initial search process yielded 16,455 potential articles, dissertations, and reports. Duplicate studies were screened out prior to evaluation for inclusion. If the same sample was used in multiple studies, then the most comprehensive evaluation was included. To be considered for inclusion in the final sample, studies had to meet the following criteria: (1) A formal mentoring program, with mentoring defined as a non-parental adult or older youth acting in a non-professional helping capacity with a specific younger person to promote positive youth outcomes through the relationship. Relationships that were more professional in nature, such as tutoring or coaching, were not included. (2) An evaluation with a comparison group, including randomized controlled trials and/or quasi-experimental studies. Most of the articles found in the initial search did not meet these inclusion criteria, due to the fact that they were theoretical rather than empirical papers, utilized qualitative analyses, or were studies of specific characteristics of mentoring relationships (e.g., termination) rather than program evaluations.

Studies then underwent a second round of screening and were excluded from the meta-analysis if they met any of the following exclusion criteria: (1) similar-age peer mentoring, (2) only group mentoring, (3) adult mentees (i.e., mentees older than 18 years of age), (4) insufficient treatment versus control group differentiation (e.g., both groups received mentoring interventions, or the treatment group included a substantial proportion of participants who never received mentoring), (5) adjunctive mentoring (i.e., evaluations in which mentoring was not one of the primary or secondary components), (6) outcomes measured did not fall into one of the following broad categories: psychological, social, school, health, or cognitive, (7) insufficient information to compute an effect size, and the author could not be reached (despite repeated attempts) to supply the needed data or did not respond to requests for additional information within a specified timeframe, and (8) the study was written in a language other than English. This procedure yielded 70 studies for analysis (see Fig. 1 for an overview of study selection and Supplementary Table A for a description of all included studies).

### Study Coding Procedures

Studies were coded for participant, program, and research design characteristics by five raters. Raters adhered to a coding manual, which outlined coding procedures and codes for each characteristic. Raters with advanced statistical experience coded study outcomes. All coders attended a training led by a researcher with expertise in meta-analytic techniques (i.e., over a decade of experience with conducting and writing about meta-analyses in the social sciences). Moreover, throughout the outcome coding process, weekly meetings were conducted in which all raters, as well as an expert consultant in meta-analytic technique, discussed and resolved difficulties or discrepancies in coding and effect size calculation. All studies were double-coded

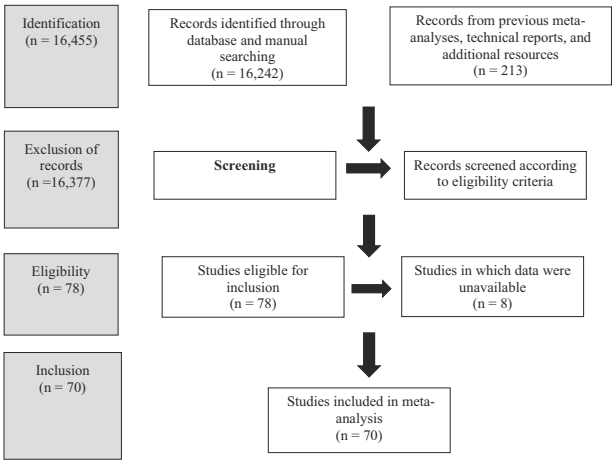


**Fig. 1** Flowchart of the study selection process. The comprehensive search for studies utilized computerized database searches (PsycINFO, ERIC, and Proquest), as well as a manual search of other resources. Studies were screened for inclusion and exclusion criteria, and authors were contacted if additional data was needed for effect size calculation. Studies for which there was insufficient data (and when authors did not respond in a specified time frame) were excluded. This procedure yielded 70 studies for analysis

and discrepancies in coding were resolved through joint review of study details and consultation of past literature. Tests of inter-rater reliability amongst the outcome codes showed that the overall agreement across coders was 92.5%.

Outcomes for each study were noted and coded for several characteristics, such as information source and statistical details (e.g., sample size, means, standard deviations, frequencies). These coded outcome characteristics were ultimately utilized to calculate an effect size for each outcome. Outcomes were coded for each study as a whole, as well as for any subgroup analyses presented within each study. Outcomes were also coded as belonging to one of the following five broad categories: psychological, social, cognitive, health, or school. These a priori categories were revisited and finalized during the coding process to ensure that they comprehensively represented the outcomes coded. Each outcome was also coded according to a more specific set of sub-categories (see Supplementary Table B for details). Finally, the type of outcome measure (i.e., questionnaire versus other), the source of outcome information (i.e., youth, parent, school, teacher, or other reporter), and the timing of the outcome assessment (i.e., immediately post-intervention versus longer term follow-up) were coded.

In addition to the outcome type, the following characteristics were coded as potential moderators of program effect sizes (see Supplementary Table B for a complete list of moderators).

### Mentee characteristics

Youth gender, age, race/ethnicity, indicators of risk, and specific target populations were examined as potential moderators of program effectiveness. Raters recorded the demographic characteristics of the sample for each study, including mentee sex, age, grade, and race/ethnicity (Hispanic/Latino, Black/African American, White, Asian, Native Hawaiian/Other Pacific Islander, American Indian/Alaska Native, Multiracial, and "other"). Several variables were coded as indicators of youth risk. First, as a proxy for low socioeconomic status, the percent of mentees receiving free or reduced-price lunch was coded for each study. Other coded indicators of risk included the percentage of mentees living in a single-parent household; percentage of mentees performing below grade level academically; and percentage of mentees with reported involvement in problem behaviors (e.g., fighting, being sent to the principal's office, suspensions, truancy, risk of dropping out of school, drug/alcohol use, early sexual activity). Finally, coders rated whether each study was designed for intervention with one of the following specific populations of youth: general population, multiple risk indicators, racial/ethnic minority youth, youth from single-parent households, youth from low-SES families, or foster care youth.

### Mentor characteristics

Mentor gender, race/ethnicity, age, student status, and involvement with a helping role or profession were examined as predictors of mentoring program effectiveness. Raters coded mentors' demographic characteristics, including sex, age, and race/ethnicity for each study. Coders also rated the percentage of mentors in the sample who were students at the time of the study (i.e., attending high school or college), and the percentage of mentors who worked in a helping profession, such as counseling or social work.

### Program characteristics

In order to examine moderation of effect sizes by program characteristics, raters coded the average number of pre-match mentor training hours for each mentoring program, as well as the expectations around program intensity (i.e., meeting frequency and length, as well as expected overall program length). In addition, raters coded whether the program was structured around program-prescribed goals (versus un-structured), and up to two of the following primary goals of the mentoring program were coded for each study: nonspecific/general positive youth development, improving academic performance, reducing behavioral

problems, reducing psychosocial problems, or a combination of academic and psychosocial goals. In addition, raters coded the program's geographical location (i.e., urban/suburban, rural, or mixed), as well as the primary program site (i.e., school versus community-based). Finally, raters coded whether or not mentors and/or mentees received any incentive for participation (e.g., course credit, payment).

### Research design and methodological factors

Finally, several aspects of each study's research design were coded in order to account for their influence on the reported effect size. The publication status (i.e., published in journal, dissertation, or report) as well as the year the study was published, defended, or presented to the public were noted. In addition, each study's design was coded as a randomized controlled trial (i.e., including a treatment condition and a no-treatment or waitlist control condition) versus a quasi-experimental design, and the control group was coded as "no treatment" versus "treatment as usual". "No treatment" was indicated for control groups that did not receive any intervention (e.g., a waitlist control), while "treatment as usual" was used for control groups that received other services offered by a program (e.g., tutoring, social services), without a specific mentoring component. Finally, a structured rating of study quality (i.e., weak, moderate, or strong) was assigned using an established procedure that accounts for study selection bias, study design, confounding variables, blinding, data collection methods, withdrawals and dropouts, intervention integrity, and analysis (National Collaborating Centre for Methods and Tools 2008).

## Effect Size Calculation and Data Analyses

The standardized mean difference between the experimental and control group was calculated as an effect size measure, with a positive value indicating an advantage for the treatment (mentoring) group over the control group. This value was transformed into Hedges' $g$ in order to adjust for differences in sample size (Hedges and Olkin 1985). When means, standard deviations, sample sizes, or other information necessary for the calculations were not reported, study authors were contacted for additional information.

In the majority of the studies, more than one effect size was calculated. A three-level approach to meta-analysis was therefore applied in order to deal with the interdependency of effect sizes (Van den Noortgate et al. 2014). The major advantage of the three-level approach is that all (dependent) effect sizes extracted from the same study can be included in the analysis, which preserves all available information. Moreover, three-level meta-analysis accounts for both within- and between-study variability, increases statistical power compared to the traditional meta-analytic approach, and facilitates the analysis of more moderators than is possible in traditional meta-analysis.

Three sources of variance are modeled in a three-level meta-analysis: the sampling variance of the observed effect sizes (level 1), the variance between effect sizes from the same study (level 2), and the variance between studies (level 3). The sampling variance of observed effect sizes (level 1) was estimated using a previously established formula (Cheung; 2014). Log-likelihood-ratio-tests were performed to compare the deviance of the full model relative to the deviance of the models excluding one of the variance parameters, which shows if significant variance is present at the second (within-study) and third (between-study) levels (Assink and Wibbelink 2016). Significant level 2 or level 3 variance indicates a heterogeneous effect size distribution, meaning that the effect sizes cannot be treated as estimates of a common effect size. In that case, moderator analyses of outcome, sample, program, and/or study methodology characteristics may explain within-study and/or between-study heterogeneity among effect sizes.

The three-level meta-analysis was conducted in R (version 3.2.0) with the metafor-package, using a multilevel random effects model (Assink and Wibbelink 2016). The restricted maximum likelihood estimate was used to estimate all model parameters, and the Knapp and Hartung (2003) method was used for testing individual regression coefficients of the meta-analytic models and for calculating the corresponding confidence intervals (see also Assink and Wibbelink 2016). Each continuous moderator was centered around its mean, and dichotomous dummy variables were created for all categorical variables (Tabachnik and Fidell 2013). In multilevel regression analyses, the intercept is the reference category, while the dummy variables test if, and to what extent, the other categories deviate from the reference category.

## Publication Bias Analyses

In meta-analysis the aim is to include all available studies previously conducted that meet the inclusion criteria (Lipsey and Wilson 2001). However, a common problem is that results may not have been reported in dissertations or research reports, or published in scientific (peer-reviewed) journals, due to non-significant or unfavorable findings and, therefore, are difficult to locate. Not including these studies may lead to an overestimation of the true effect size, a phenomenon known as "publication bias" (Rosenthal 1979).

In order to check for evidence of publication bias, the present analyses first examined differences in effect sizes between dissertations, research reports, and scientific journal articles. An alternative way of examining the potential effect of publication bias on meta-analytic results uses a funnel

**Table 1** Descriptive information for coded moderators

| Moderator | Minimum | Maximum | Mean |
|---|---|---|---|
| *Youth Characteristics* | | | |
| Percentage male | 0 | 100 | 55 |
| Age | 9 | 16 | 12 |
| Percentage White | 0 | 88 | 32 |
| Percentage Black | 0 | 100 | 43 |
| Percentage Hispanic | 0 | 90 | 26 |
| Percentage Asian | 0 | 100 | 5 |
| Percentage Hawaiian | 0 | 12 | 1 |
| Percentage American Indian | 0 | 5 | 1 |
| Percentage multiethnic | 0 | 33 | 4 |
| Percentage other ethnicity | 0 | 35 | 13 |
| Percentage single parent household | 25 | 100 | 63 |
| Percentage free lunch (SES) | 17 | 100 | 72 |
| Percentage below grade academic functioning | 51 | 100 | 82 |
| Percentage problem behavior | 12 | 100 | 83 |
| *Mentor Characteristics* | | | |
| Percentage Male | 0 | 100 | 42 |
| Age | 20 | 57 | 36 |
| Percentage White | 0 | 100 | 62 |
| Percentage Black | 0 | 100 | 31 |
| Percentage Hispanic | 0 | 54 | 9 |
| Percentage Asian | 0 | 100 | 6 |
| Percentage Hawaiian | 0 | 1 | 0 |
| Percentage American Indian | 0 | 7 | 1 |
| Percentage multiethnic | 0 | 6 | 1 |
| Percentage other ethnicity | 0 | 19 | 10 |
| Percentage helping professionals | 0 | 100 | 79 |
| Percentage student mentors | 0 | 100 | 48 |
| *Program Characteristics* | | | |
| Pre-match training hours | 1 | 16 | 4 |
| Program length in months | 2 | 60 | 11 |
| Meeting length in hours | 0.5 | 4 | 1.7 |
| Meeting frequency | 1 | 20 | 4 |
| Program location | 87% urban/suburban, 4% rural, 9% mixed | | |
| Primary site | 63% school-based, 37% community-based | | |
| Curriculum-based | 12% yes, 88% no | | |
| Program structure | 62% unstructured, 21% semi-structured, 17% structured | | |
| Mentor incentive | 39% yes, 61% no | | |
| Mentee incentive | 35% yes, 65% no | | |

plot to inspect the distribution of points when each individual study's effect size (on the horizontal axis) is plotted against its precision, indicated by the reciprocal of the standard error (on the vertical axis). In the absence of publication bias, the distribution of plotted points should be shaped as a funnel, since the studies with small sample sizes are expected to show a larger variation in the magnitude of effect sizes, given the relatively large standard errors, than the studies with large sample sizes, with relatively small standard errors. A violation of funnel plot symmetry reflects potential publication bias, or a selective inclusion of studies showing positive or negative outcomes (Sutton et al. 2000). The effect of funnel plot asymmetry on the magnitude of the observed effect size can be examined by means of trim and fill procedures, which involve removing the asymmetric right- or left-hand side of the funnel in order to estimate the true center of the funnel, and then subsequently replacing the trimmed studies and their counterparts around the center.

The present study used both a funnel plot and a trim-and-fill analysis (Duval and Tweedie 2000), conducted with the function 'trimfill' in the metafor package (Viechtbauer 2010). All effect sizes were aggregated at the publication level (because publication bias is a publication-level phenomenon). Subsequently, trim and fill analyses tested for publication bias by examining whether effect sizes were missing on the left side of the distribution of effect sizes (indicating missing statistically non-significant or negative results). In contrast, missing effect sizes at the right side of the funnel would indicate selection bias due to an over-representation of studies with particular characteristics that might be systematically associated with *larger* effect sizes.

## Results

### Average Effect of Mentoring

There were 70 studies providing estimates of effect sizes of the impact of youth mentoring, with a combined sample size of 25,286 mentees. The average effect size across all 70 studies and all outcomes was $\overline{g} = .21$ ($p < .001$; 95% CI: .14–.28). The analyses revealed that there was significant heterogeneity across studies ($\sigma^2_{\text{level 3}} = .07$, $p < .001$), as well as significant variability between effect sizes extracted from the same study ($\sigma^2_{\text{level 2}} = .04$, $p < .001$). Notably, 33% of the variance among effect sizes was accounted for by the within-study level, and 55% by the between-study level, while random sampling error accounted for 12% of the variance. To explore this substantial variability, a number of moderators were considered as predictors of variability in youth mentoring effects both across and within studies (descriptive statistics for coded moderators are presented in Table 1).

### Differences in Effects of Mentoring Based on Youth Outcome Type

Analysis of within-study, assessment-specific characteristics that might have accounted for heterogeneity across effect sizes are presented in Table 2. Many studies reported effect sizes for mentoring programs across a range of diverse youth outcomes, such as depressive symptoms, school grade point average, or parent-child relationship quality. Using a multi-level approach, analyses first explored whether the effects of youth mentoring were different across five broad outcome categories: school, social, health,

**Table 2** Within-study moderators of the effectiveness of mentoring programs: Assessment of outcomes

| Moderator variable | *k* | *#ES* | $B_0$/ g | $t_0$ | $B_1$ | $t_1$ | $F(df_1, df_2)$ |
|---|---|---|---|---|---|---|---|
| *Broad outcome domains* | | | | | | | $F(4, 690) = 0.68$ |
| School (RC) | 51 | 222 | 0.20 | $6.27^{***}$ | | | |
| Psychological outcomes | 46 | 191 | 0.17 | $5.01^{***}$ | −0.03 | −1.23 | |
| Health | 17 | 38 | 0.23 | $4.76^{***}$ | 0.03 | 0.61 | |
| Cognitive functioning | 35 | 80 | 0.19 | $4.81^{***}$ | −0.01 | −0.39 | |
| Social | 43 | 164 | 0.19 | $5.82^{***}$ | −0.01 | −0.18 | |
| *Outcome Sub-Categories* | | | | | | | |
| School | | | | | | | $F(2, 219) = 2.49^{+}$ |
| Academic outcomes (RC) | 38 | 115 | 0.16 | $3.58^{***}$ | | | |
| School engagement | 44 | 102 | 0.25 | $5.53^{***}$ | 0.08 | $2.06^{*}$ | |
| Extracurricular activities | 3 | 5 | 0.12 | 0.96 | −0.04 | −0.36 | |
| Psychological symptoms | | | | | | | $F(3, 187) = 0.56$ |
| Externalizing (RC) | 38 | 106 | 0.15 | $3.72^{***}$ | | | |
| Internalizing | 20 | 49 | 0.18 | $3.47^{***}$ | 0.03 | 0.66 | |
| Other mental health | 6 | 20 | 0.10 | 1.19 | −0.05 | −0.59 | |
| Self-regulation | 9 | 16 | 0.22 | $2.48^{*}$ | 0.07 | 0.77 | |
| Health | | | | | | | $F(2, 35) = 1.29$ |
| Substance use (RC) | 11 | 27 | 0.09 | $1.85^{+}$ | | | |
| Physical health | 4 | 5 | 0.27 | $2.14^{*}$ | 0.18 | 1.35 | |
| Well-being | 4 | 6 | 0.21 | $1.91^{+}$ | 0.12 | 1.01 | |
| Cognition | | | | | | | $F(1, 78) = 0.32$ |
| Executive functioning (RC) | 6 | 11 | 0.18 | $2.65^{**}$ | | | |
| Self-cognition | 34 | 69 | 0.14 | $5.07^{***}$ | −0.04 | −0.56 | |
| Social functioning | | | | | | | $F(2, 161) = 3.50^{*}$ |
| Social skills | 20 | 50 | 0.07 | 1.59 | | | |
| Social support | 14 | 22 | 0.20 | $3.61^{***}$ | 0.13 | $2.14^{*}$ | |
| Relationships | 33 | 92 | 0.18 | $4.64^{***}$ | 0.11 | $2.37^{*}$ | |
| *Approaches to Outcome Measurement* | | | | | | | |
| Type of Measure | | | | | | | $F(1, 693) = 5.68^{*}$ |
| Questionnaires | 57 | 532 | 0.22 | $6.94^{***}$ | | | |
| Other Measures | 36 | 163 | 0.14 | $3.72^{***}$ | −0.08 | $-2.38^{*}$ | |
| Information source | | | | | | | $F(4, 690) = 2.58^{*}$ |
| Youth | 55 | 421 | 0.22 | $6.96^{***}$ | | | |
| Parent | 11 | 61 | 0.22 | $4.31^{***}$ | 0.00 | 0.09 | |
| School | 28 | 125 | 0.12 | 2.96** | −0.09 | $-2.61^{**}$ | |
| Teacher | 13 | 61 | 0.14 | $2.90^{**}$ | −0.08 | $-2.17^{*}$ | |
| Other | 8 | 27 | 0.21 | $2.84^{**}$ | −0.01 | −0.17 | |
| Time | | | | | | | $F(1, 693) = 1.97$ |
| Post-test | 69 | 604 | 0.19 | $6.51^{***}$ | | | |
| Follow-up | 11 | 91 | 0.24 | $5.55^{***}$ | 0.49 | 1.40 | |

*RC* reference category, *k* number of independent studies, *#ES* number of effect sizes, $B_0/g$ intercept/mean effect size, $t_0$ difference in mean effect size with zero, $B_1$ estimated regression coefficient, $t_1$ difference in mean effect size with RC, $F(df_1, df_2)$ omnibus test

$^{+}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

cognitive, and psychological outcomes. Results suggested no significant differences in effect sizes across these five types of outcomes ($F(4, 700) = .31$, $p = .78$), with effect sizes for all outcome categories significantly different from 0 (all $t$'s > 4.24, all $p$'s < .001).

Analyses next evaluated an even more fine-grained coding of outcome types. Specifically, analyses compared effect sizes

**Table 3** Between-study moderators of the effectiveness of mentoring programs

| Moderator variable | *k* | *#ES* | $B_0$/g | $t_0$ | $B_1$ | $t_1$ | $F(df_1, df_2)$ |
|---|---|---|---|---|---|---|---|
| *Mentee characteristics* | | | | | | | |
| Percentage male | 58 | 605 | 0.21 | $5.62^{***}$ | 0.38 | $2.19^{*}$ | $F(1, 603) = 4.81^{*}$ |
| Age | 63 | 645 | 0.21 | $5.62^{***}$ | 0.02 | $1.67^{+}$ | $F(1, 643) = 2.78^{+}$ |
| Ethnicity | | | | | | | |
| Percentage White | 48 | 514 | 0.23 | $4.80^{***}$ | 0.00 | 0.02 | $F(1, 512) = 0.00$ |
| Percentage Black | 46 | 490 | 0.21 | $5.53^{***}$ | 0.01 | 0.10 | $F(1, 488) = 0.01$ |
| Percentage Hispanic | 44 | 478 | 0.24 | $4.77^{***}$ | 0.14 | 0.66 | $F(1, 476) = 0.44$ |
| Percentage Asian | 39 | 435 | 0.24 | $4.36^{***}$ | −0.30 | −0.89 | $F(1, 433) = 0.80$ |
| Percentage Hawaiian | 35 | 401 | 0.21 | $4.08^{***}$ | 3.84 | $1.86^{+}$ | $F(1, 399) = 2.83^{+}$ |
| Percentage Indian | 38 | 433 | 0.23 | $4.56^{***}$ | 2.02 | 0.73 | $F(1, 431) = 0.53$ |
| Percentage multi-ethnic | 30 | 336 | 0.26 | $4.10^{***}$ | −0.04 | −0.05 | $F(1, 334) = 0.00$ |
| Percentage other ethnicity | 30 | 342 | 0.24 | $3.69^{***}$ | −0.26 | −0.78 | $F(1, 340) = 0.61$ |
| Risk indicators | | | | | | | |
| Percentage single parent household | 20 | 270 | 0.15 | $3.87^{***}$ | 0.05 | 0.30 | $F(1, 268) = 0.09$ |
| Percentage free lunch (SES) | 13 | 126 | 0.18 | $2.69^{**}$ | 0.27 | 0.86 | $F(1, 124) = 0.74$ |
| Percentage below grade academic functioning | 7 | 61 | 0.24 | $3.26^{**}$ | 0.53 | 1.60 | $F(1, 59) = 2.57$ |
| Percentage problem behavior | 14 | 131 | 0.18 | $1.85^{+}$ | 0.33 | 1.39 | $F(1129) = 1.93$ |
| Type of sample | | | | | | | $F(5, 699) = 0.10$ |
| General population | 15 | 138 | 0.17 | $2.27^{*}$ | | | |
| Multiple risk | 34 | 321 | 0.22 | $4.46^{***}$ | 0.05 | 0.58 | |
| Minority | 9 | 81 | 0.21 | $2.14^{*}$ | 0.04 | 0.33 | |
| Single parent | 4 | 41 | 0.17 | 1.17 | 0.00 | 0.02 | |
| Low SES | 6 | 92 | 0.23 | $2.04^{*}$ | 0.06 | 0.46 | |
| Foster care | 2 | 32 | 0.18 | 0.89 | 0.01 | 0.04 | |
| *Mentor characteristics* | | | | | | | |
| Percentage male | 33 | 410 | 0.20 | $4.19^{***}$ | 0.36 | $2.14^{*}$ | $F(1, 408) = 4.56^{*}$ |
| Age | 11 | 189 | 0.16 | $2.78^{**}$ | 0.01 | 1.41 | $F(1187) = 1.98$ |
| Ethnicity | | | | | | | |
| Percentage White | 24 | 297 | 0.24 | $3.61^{***}$ | −0.10 | - 0.50 | $F(1, 295) = 0.25$ |
| Percentage Black | 23 | 268 | 0.25 | $3.74^{***}$ | 0.24 | 1.09 | $F(1, 266) = 1.18$ |
| Percentage Hispanic | 22 | 258 | 0.25 | $3.38^{***}$ | −0.09 | - 0.19 | $F(1, 256) = 0.04$ |
| Percentage Asian | 22 | 254 | 0.27 | $3.84^{***}$ | −0.43 | −1.29 | $F(1, 252) = 1.67$ |
| Percentage Hawaiian | 22 | 258 | 0.26 | $3.41^{***}$ | −19.36 | −0.58 | $F(1, 256) = 0.33$ |
| Percentage American Indian | 22 | 258 | 0.27 | $3.59^{***}$ | −5.87 | −0.58 | $F(1, 256) = 0.33$ |
| Percentage multi-ethnic | 22 | 258 | 0.28 | $3.73^{***}$ | −4.21 | −0.77 | $F(1, 256) = 0.59$ |
| Percentage other ethnicity | 23 | 299 | 0.26 | $3.62^{***}$ | −0.08 | −0.24 | $F(1, 297) = 0.81$ |
| Percentage helping professionals | 31 | 291 | 0.19 | $3.69^{***}$ | 0.25 | $2.34^{*}$ | $F(1, 289) = 5.49^{*}$ |
| Percentage student mentors | 32 | 314 | 0.16 | $3.68^{***}$ | −0.15 | $-1.68^{+}$ | $F(1, 312) = 2.82^{+}$ |
| *Program Characteristics* | | | | | | | |
| Pre-match training hours | 20 | 198 | 0.34 | $3.40^{***}$ | 0.01 | 0.27 | $F(1, 196) = 0.07$ |
| Program length in months | 30 | 331 | 0.24 | $3.82^{***}$ | 0.00 | 0.15 | $F(1, 329) = 0.02$ |
| Meeting length in hours | 31 | 348 | 0.17 | $4.66^{***}$ | −0.01 | $-1.98^{*}$ | $F(1, 346) = 3.92^{*}$ |
| Meeting frequency | 44 | 459 | 0.22 | $4.51^{***}$ | −0.01 | −0.54 | $F(1, 457) = 0.30$ |
| Primary focus | | | | | | | $F(4, 700) = 1.08$ |
| General | 29 | 336 | 0.14 | $2.63^{**}$ | | | |

**Table 3** (continued)

| Moderator variable | *k* | *#ES* | $B_0$/g | $t_0$ | $B_1$ | $t_1$ | $F(df_1, df_2)$ |
|---|---|---|---|---|---|---|---|
| Academic | 17 | 133 | 0.30 | $4.41^{***}$ | 0.16 | $1.96^{+}$ | |
| Behavior problems | 13 | 82 | 0.25 | $3.04^{**}$ | 0.11 | 1.17 | |
| Psychosocial problems | 5 | 57 | 0.19 | 1.49 | 0.05 | 0.39 | |
| Academic and psychosocial | 6 | 97 | 0.25 | $2.23^{*}$ | 0.11 | 0.91 | |
| Geographical location | | | | | | | $F(2, 451) = 0.62$ |
| Urban & Suburban | 39 | 393 | 0.24 | $5.02^{***}$ | | | |
| Rural | 3 | 19 | 0.32 | $1.77^{+}$ | 0.08 | 0.42 | |
| Mixed | 4 | 46 | 0.09 | 0.61 | −0.15 | −1.00 | |
| Primary Site | | | | | | | $F(1, 560) = 3.44^{+}$ |
| School | 33 | 314 | 0.27 | $5.22^{***}$ | | | |
| Community | 19 | 248 | 0.12 | $1.75^{+}$ | −0.16 | $-1.85^{+}$ | |
| Curriculum based | | | | | | | $F(1, 651) = 0.03$ |
| Yes | 8 | 73 | 0.24 | $2.29^{*}$ | | | |
| No | 52 | 580 | 0.22 | $5.47^{***}$ | −0.02 | −0.16 | |
| Program Structure | | | | | | | $F(2, 683) = 0.72$ |
| Unstructured | 43 | 446 | 0.25 | $5.54^{***}$ | | | |
| Semi-structured | 14 | 149 | 0.14 | $1.86^{+}$ | −0.10 | −1.18 | |
| Structured | 10 | 91 | 0.20 | $2.17^{*}$ | −0.05 | −0.47 | |
| Mentor incentive | | | | | | | $F(1, 486) = 0.03$ |
| Yes | 20 | 211 | 0.22 | $3.24^{**}$ | | | |
| No | 30 | 277 | 0.20 | $3.66^{***}$ | −0.02 | −0.19 | |
| Mentee incentive | | | | | | | $F(1, 469) = 1.02$ |
| Yes | 15 | 199 | 0.29 | $3.79^{***}$ | | | |
| No | 35 | 272 | 0.20 | $3.77^{***}$ | −0.09 | −1.01 | |
| *Methodological Characteristics* | | | | | | | |
| Year of publication | 70 | 695 | 0.20 | $6.05^{***}$ | −0.01 | −1.18 | $F(1, 703) = 1.40$ |
| Publication status | | | | | | | $F(2, 702) = 1.01$ |
| Published (journal) | 35 | 341 | 0.19 | $4.01^{***}$ | | | |
| Dissertation | 24 | 198 | 0.27 | $4.61^{***}$ | 0.08 | 1.01 | |
| Report | 11 | 166 | 0.14 | $1.68^{+}$ | −0.06 | −0.62 | |
| Study Design | | | | | | | $F(1, 703) = 0.44$ |
| RCT | 38 | 455 | 0.19 | $4.16^{***}$ | | | |
| Quasi experimental | 32 | 250 | 0.23 | $4.61^{***}$ | 0.04 | 0.66 | |
| Type of control group | | | | | | | $F(1, 703) = 2.00$ |
| No treatment | 62 | 597 | 0.23 | $6.31^{***}$ | | | |
| Treatment As Usual (TAU) | 8 | 108 | 0.08 | 0.78 | −0.15 | −1.41 | |
| Study quality | | | | | | | $F(2, 702) = 0.06$ |
| Strong | 17 | 249 | 0.19 | $2.90^{**}$ | | | |
| Moderate | 23 | 255 | 0.20 | $3.34^{***}$ | 0.01 | 0.10 | |
| Weak | 30 | 201 | 0.22 | $4.19^{***}$ | 0.03 | 0.33 | |

*RC* reference category *k* number of studies, *#ES* number of effect sizes, $B_0/g$ intercept/ mean effect size, $t_0$ difference in mean effect size and zero, $B_1$ estimated regression coefficient, $t_1$ difference in mean effect size with RC, $F(df_1, df_2)$ omnibus test

$^{+}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

of mentoring on 15 types of sub-categories, nested within the broader outcome types. An omnibus test revealed no substantial variability across all 15 outcome sub-types ($F(14, 690) = 0.77$, $p = .70$). Consistent with this result, there was no significant variability among the psychological outcomes (externalizing symptoms, internalizing symptoms, self-regulation problems, and other mental heath measures), health outcomes (substance use, physical health, and general

well-being), or cognitive outcomes (executive functioning and self-cognition). Larger effects were observed for mentoring programs on measures of social support ($B = .20$, $t = 3.29$, $p < .01$) and relationship quality ($B = .18$, $t = 4.46$, $p < .001$) than on measures of social skills ($B = .08$, $t = 1.78$, $p < .10$), although it should be noted that the overall test of variability in social outcome types did not reach statistical significance ($F(2, 168) = 2.79$, $p < .10$).

Finally, analyses also examined whether the type of outcome measure (questionnaire versus other methods), the source of outcome information (youth, parent, school, teacher, or other report), and the timing of the outcome assessment (immediately post-intervention versus longer term follow-up) accounted for variability in effect sizes. Results showed that there was significant variability in effect sizes across assessment types ($F(1, 703) = 4.47$, $p < .05$), with questionnaire measures yielding larger effect sizes than other assessment types, such as interviews or examination of school records ($B = -.08$, $t = -2.11$, $p < .05$). In addition, both school records ($B = -.10$, $t = -2.38$, $p < .05$) and teacher reports ($B = -.10$, $t = -2.16$, $p < .05$) yielded smaller effect sizes than youth self-report (no differences noted for parent or other informants, relative to youth self-report), although again it should be noted that the test of overall variability in effect sizes across informants did not reach statistical significance ($F(4, 700) = 2.30$, $p < .10$). There were no differences in effect sizes for assessments immediately post-intervention versus at a later follow-up ($F(1, 703) = 2.48$, $p = .12$).

## Between-Study Moderators of Mentoring Effectiveness

Results of moderator analyses using between-study youth, mentor, and program characteristics are summarized in Table 3.

### Youth characteristics

Results showed that there were statistically significant differences in the impact of youth mentoring based on the percentage of male youth within the sample ($F(1, 603) = 4.81$, $p < .05$), with larger effects in samples that had a higher percentage of male youth ($B = .38$, $t = 2.19$, $p < .05$). There were no differences in study effect sizes as a function of youth age or youth race/ethnicity, and there were also no differences in effect sizes based on indicators of youth risk at baseline, including the percentage of single-parent households, youth receiving free or reduced-price lunch, youth performing below grade level, and youth exhibiting problem behaviors. Mentoring programs designed for specific populations of youth, such as youth in foster care or minority youth, also showed no significant differences in effects, relative to mentoring programs designed for a general population of youth.

### Mentor characteristics

Results showed that there were significant differences in the impact of youth mentoring based on the percentage of male mentors within the sample ($F(1, 408) = 4.56$, $p < .05$), with larger effects in samples that had a higher percentage of male mentors ($B = .36$, $t = 2.14$, $p < .05$). Samples with a higher percentage of mentors who worked within the helping professions also showed higher effect sizes for youth outcomes ($F(1, 289) = 5.49$, $p < .05$; $B = .25$, $t = 2.34$, $p < .05$). There were no differences in effect size based on the breakdown of mentor ethnic background or student status, nor did average mentor age moderate program effects.

### Program characteristics

Expectations around match meeting duration predicted variability in program effect sizes ($F(1, 346) = 3.92$, $p < .05$), with programs that specified longer meeting durations for matches tending to yield *smaller* effect sizes ($B = -.01$, $t = -1.98$, $p < .05$). There was no observed effect of program site (community-based versus school-based) on mentoring impact, and there were also no differences in the impact of youth mentoring programs based on program length, program expectations around match meeting frequency, average number of pre-match training hours for mentors, or provision of mentor and/or youth incentives for program participation. Likewise, no moderation was observed for the geographical location of the program, whether the program was more structured, or the primary focus (e.g., general versus psychosocial versus academic) of the program.

## Methodological Predictors of Mentoring Effect Sizes

The findings from analyses of methodological characteristics as predictors of effect size are summarized in Table 3. There were no significant differences in effect sizes based on year of study publication or publication status of the study (i.e., published in a journal versus dissertation or research report). Likewise, there were no differences in the reported impact of mentoring based on the specific study design (randomized controlled trial versus quasi-experimental), control group utilized (no treatment versus "treatment as usual"), or ratings of overall study quality.

## Publication Bias Analyses

A funnel plot analysis revealed that publication bias was unlikely. No studies were missing at the left side of the funnel plot (see Fig. 2), and examining the distribution of

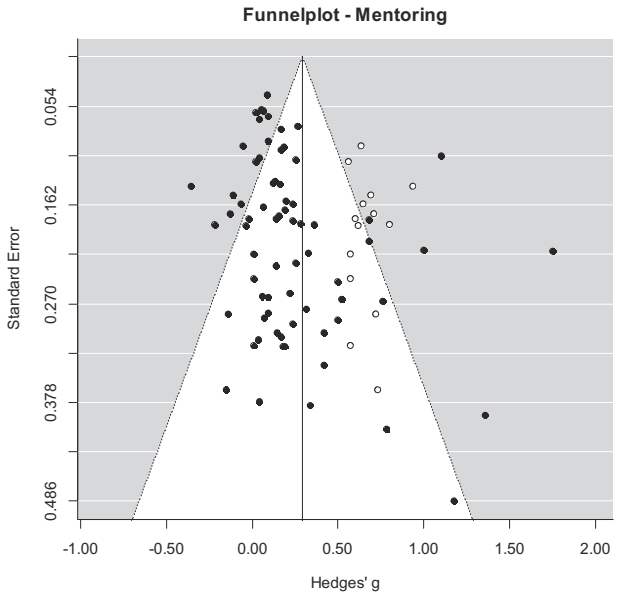


**Fig. 2** Funnel plot testing for publication bias analysis. Data were plotted to examine the potential for bias caused by un-published results not included in the current meta-analysis. The funnel plot, showing no studies missing at the left side of the funnel, suggests that publication bias was unlikely. However, 14 effect sizes were missing at the right side of the funnel plot, which suggests a potential selection bias

effect sizes at the study level yielded an overall mean effect size of Hedges' $g = 0.21$ ($p < .001$), which did not differ from the overall mean effect size obtained by the three-level meta-analytic procedure (Hedges' $g = 0.21$, $p < .001$). However, 14 effect sizes were missing at the right side of the of the funnel plot, which suggests a possible selection bias. That is, the plot provides some indication that the present analyses might have excluded studies with *larger* effect sizes, thereby under-estimating the overall effect of youth mentoring. Accounting for selection bias by means of a trim and fill analysis yielded a somewhat larger mean effect size of Hedges' $g = 0.29$ ($p < .001$).

## Discussion

The last comprehensive meta-analysis of youth mentoring programs included outcome studies through 2010 (Dubois et al. 2011); however, since that time, there has been a proliferation of large-scale evaluations of mentoring interventions, in line with an increased focus on evidence-based guidelines for mentoring youth (e.g., the establishment of a National Mentoring Resource Center for supporting evidence-based mentoring practices). The present study involved a comprehensive meta-analysis of all youth mentoring program evaluations written in the English language through 2017. A multilevel meta-analytic approach was used to estimate the overall effect size of youth mentoring programs, as well as explore predictors of within- and between-study variability in effect sizes. In addition, the current study aimed to improve on past meta-analytic findings by using selection criteria that adhered more strictly to a relationship-focused developmental model of mentoring, and therefore included only those studies that evaluated intergenerational, one-on-one youth mentoring programs (excluding, for example, studies of tutoring or peer mentoring).

### Overall Effects of Youth Mentoring

Analyses of the data from 70 studies of youth mentoring programs revealed that the mean effect of mentoring on youth outcomes was .21. Although this effect is considered small by Cohen's (1988) guidelines, it falls well within the medium/moderate range of empirical guidelines for the average effect sizes of universal youth prevention programs (Tanner-Smith et al. 2018). At the same time, however, many youth who are referred to mentoring programs are already experiencing sub-clinical levels of difficulties and symptoms, and thus present greater room for improvement on outcome assessments than youth in primary prevention programs (Jarjoura et al. 2018). As such, comparisons with the somewhat larger effects reported in indicated (secondary) prevention programs may also be warranted (Durlak and Wells 1998; Tanner-Smith et al. 2018).

The effect size observed in these analyses is remarkably consistent with past comprehensive meta-analyses of youth mentoring, which have shown overall effect sizes ranging from .18 to .21 (DuBois et al. 2002; 2011). The consistency over time and across meta-analytic study designs is particularly notable given several methodological differences in the present study compared to previous studies. First, this sample of studies included only those mentoring programs that were designed to improve youth outcomes through a one-to-one mentoring relationship with a caring adult (rather than a purely curriculum-based, peer, or group mentoring approach). This stricter definition was used in order to identify mentoring programs that were most consistent with the prevailing theory of the developmental mechanisms hypothesized to mediate the impact of mentoring on youth outcomes (Rhodes 2005).

The consistency in overall effects across studies is also notable given the inclusion in the present meta-analysis of more recent evaluations. In the past decade, there has been an increase in mentoring programs that implement evidence-based program practices, rather than relying solely on practice wisdom. Recent increases in advocacy and support from organizations like MENTOR, as well as the publication of the Elements of Effective Practice for Mentoring (EEPM; Garringer et al. 2015) have been initiated in an effort to increase awareness of the link between

empirically-based practices and mentoring effectiveness among practitioners. In fact, greater implementation of the benchmark practices defined in the EEPM has been linked with match longevity (Kupersmidt et al. 2017b), a key predictor of stronger mentoring outcomes in previous studies (Grossman et al. 2012). It appears that these improvements in relationship quality and longevity may not have translated into substantially larger effects on more distal youth outcomes. Nevertheless, it is important to note that even small to moderate improvements in aspects of youth functioning like substance use, depressive symptoms, and academic engagement can have an important influence on trajectories of positive youth development, especially when these improvements occur during critical periods of development (Tanner-Smith et al. 2018).

One unique feature of the present study involved the use of multi-level meta-analyses to account for variability in effect sizes both within and across studies. The analyses suggested that a substantial portion of the variance among effect sizes was accounted for within studies (33%). That is, even within the same program evaluation, there were significant differences in the size of observed effect sizes for different outcome measures or constructs. This finding suggests that the multi-level approach to meta-analysis should be considered as a particular strength of this study, given that a traditional meta-analysis only accounts for between-study heterogeneity among effect sizes. Future studies in this area should use a multi-level approach to more accurately explore the specific outcome-level (e.g., self-report versus teacher-report) and study-level (e.g., demographic background of the youth in the program) constructs that might be accounting for all sources of variability in effect sizes.

## Differences in Effects of Mentoring Based on Youth Outcome Type

Statistically significant effects were observed across all outcome domains assessed, including school, cognitive, health, psychological, and social outcomes, and no significant difference in effect size across these broad domains of functioning was detected. Effect sizes were also largely similar across the narrower sub-categories of outcomes. There were some noted differences among various youth social outcomes, with mentoring programs showing larger effects on youth perceptions of social support, as well as youth reports of the quality of their relationships with peers, teachers, and parents, relative to youth social skills. These findings are consistent with a developmental framework of mentoring, which posits that a relationship with a caring adult influences children's cognitive representations of relationships (Bowlby 1988), thereby enhancing a youth's perceptions of social support (Rhodes 2005). Mentees' improved perceptions of support, in turn, may lead to improvements in a wide range of developmentally-relevant outcomes, including academic engagement and performance, self-esteem, assertiveness, and substance use (Chan et al. 2013; Karcher et al. 2002).

Although these results are generally similar to those reported in previous meta-analyses, in that effects of mentoring were observed across a wide range of outcomes, there are some notable differences. For example, some meta-analyses have shown relatively low effect sizes for psychological/emotional outcomes (DuBois et al. 2002, 2011), but relatively strong effects of mentoring on academic problems (DuBois et al. 2011). In contrast, the current meta-analysis excluded academic tutoring that relied on highly structured curricula, which may have attenuated academic outcomes in these analyses. This finding suggests that relationship-oriented mentoring programs need not be limited to youth in need of academic support. Instead, mentoring appears to be a modestly effective strategy for promoting a wide range of positive outcomes, including mental health, early in development.

Additionally, DuBois and colleagues' (2011) meta-analysis found an effect size of close to zero for physical health outcomes, whereas the current meta-analysis showed an effect size of .24. Again, this difference in magnitude is likely due to the two meta-analyses employing different definitions of health outcomes, with the DuBois meta-analysis focusing primarily on repeat pregnancies and body mass index, while the current meta-analysis defined health more broadly, including substance use, physical health, and general well-being. Notably, however, even within the narrow domain of physical health outcomes, a relatively large effect size of .27 was observed. Very few evaluations included in the meta-analysis have measured physical health outcomes, another factor that might account for variability in findings across studies. More generally, the differences in results may be explained by the fact that the present analyses took a novel approach to exploring various domains of youth functioning in two ways: first, a more fine-grained coding scheme was developed to map onto key constructs within clinical and developmental science; and multi-level meta-analyses were used in order to fully address issues of within-study variability across outcome domains.

These findings highlight several key issues for research and practice. First, although some programs have begun to embrace more targeted, evidence-based approaches, mirroring recent trends within the practice of psychotherapy, mentoring programs, overall, appear to be producing small- to moderate-sized effects across a broad range of outcomes, and reported results tend to be inconsistent across evaluations. These effect sizes may grow if the field continues to embrace more rigorous adherence to evidence-based practices that target specific mechanisms underlying particular

youth difficulties, rather than relying on a relatively low-intensity, nonspecific approach with uneven adherence to practices that are research-informed. Moreover, there may be benefits to aligning assessment of youth outcomes with clinical, developmental, and prevention science rather than measuring outcomes as broad, heterogeneous categories (e.g., "psychosocial" outcomes) that can obscure the specificity of mentoring program effects.

Another issue involves the methods used for choosing assessments of youth outcomes in mentoring programs. The most commonly used method of assessment within mentoring programs, questionnaires, yielded significantly larger effect sizes than other assessment methods, such as interviews or school records. In addition, both school records and teacher reports yielded lower effect sizes than youth self-report measures. These findings are consistent with meta-analyses in other fields, which have highlighted the fact that assessment approaches can substantially impact the observed impact of a treatment, particularly when outcome measures are developed by the evaluators themselves (Cheung and Slavin 2015). Mentoring researchers and practitioners should be aware that the types of assessments they choose, regardless of the construct being measured, could influence their evaluation of the mentoring program's effectiveness. Interestingly, there were no differences in effect sizes for assessments administered immediately post-intervention versus those administered at a later follow-up. Although this may seem surprising in the context of individual studies of mentoring programs that have shown erosion of effects over time (e.g., Herrera et al. 2011), it may be that the studies that included follow-up assessments were conducted on stronger, longer-lasting programs. Additionally, it should be noted that only 11 studies included follow-up assessments, calling attention to the need for evaluations to include longer-term follow-ups.

## Moderators of the Effectiveness of Youth Mentoring

The current meta-analysis also tested several potential moderators of youth mentoring effectiveness in an effort to explain the substantial heterogeneity in effect sizes across studies and programs. Larger effects were observed for programs that served a higher percentage of male youth, consistent with at least one previous meta-analysis of mentoring, which showed that programs serving more than 50% males had stronger effects (DuBois et al. 2011). Girls may enter mentoring programs with more complicated relational histories than boys, which may initially hamper mentors' capacity to forge productive ties with them (Bogat and Liang 2005) and lead to premature closure (Kupersmidt et al. 2017a). Mentoring also had a larger impact in samples that had a higher percentage of male mentors; however, because many programs attempt to match mentor and youth gender, it is difficult to tease apart the effects of youth versus mentor gender in these findings. Of course, since the comparisons were made on the basis of program gender composition, as opposed to direct comparisons between males and females, the differences may have more to do with unspecified program or study features. Further research is therefore needed to fully untangle the ways in which mentor and youth gender might influence match outcomes.

Consistent with past literature (DuBois et al. 2002, 2011), the effectiveness of mentoring did not appear to vary substantially depending on mentee race or ethnicity. However, it is somewhat surprising that there were not different effects of mentoring based on youth age. Past research tends to show that older youth have less close and enduring mentoring relationships (e.g., Herrera et al. 2000; Kupersmidt et al. 2017a). Further research is warranted to determine whether, despite potentially poorer relationship processes, mentoring relationships can still be equally influential for older youth.

Also consistent with previous studies, programs with a greater percentage of mentors who worked in helping professions showed larger effect sizes for youth outcomes (DuBois et al. 2002). Volunteers who have had previous experience with helping youth may feel a stronger sense of efficacy, a variable that has been consistently associated with better match outcomes (e.g., Karcher et al. 2005). For example, one recent study found that volunteer mentors with greater self-efficacy and more previous involvement with youth in their communities were more successful in working with youth from high-stress backgrounds than mentors with lower self-efficacy and less previous experience (Raposa et al. 2016). This finding underscores the benefits of recruiting mentors with helping experiences or roles, and of providing less experienced volunteers with adequate training prior to the initiation of the match. There was no evidence of substantially lower effect sizes for high school or college student mentors, in contrast to previous evaluations (e.g., Herrera et al. 2011). Future research should attempt to clarify this discrepancy, given that student mentors are especially essential to the success of school-based mentoring programs, which have grown in popularity in recent years.

With growing interest in mentoring programs, and substantial diversity in the practice of mentoring, it is becoming increasingly important to examine how program structure and mentoring approaches influence the effectiveness of programs. The lack of significant differences in effect sizes based on program length is inconsistent with previous research and theory indicating the importance of longer relationships (e.g., Grossman and Rhodes 2002); however, program length was also not a significant predictor of outcomes in earlier meta-analyses (DuBois et al. 2002, 2011). This finding suggests that, although relationship length may

be an important determinant of youth outcomes *within* a specific relationship or program, it may be less important in distinguishing overall levels of mentoring effectiveness *across* programs. In fact, provided that there is sufficient opportunity to forge a productive bond, certain targeted, time-limited mentoring programs have been shown to yield positive youth outcomes (Cavell and Elledge 2014; Taussig and Culhane 2010). Additionally, as noted, fulfilling length commitments is more important than the actual length of the relationship (Grossman et al. 2012). Thus, further exploration of these issues with more precise and comprehensive measurement is warranted to determine how long it takes for the results of mentoring to appear, whether shorter, targeted programs are more effective, and whether predictors of duration vary across subgroups of youth and mentors.

Interestingly, programs that had expectations for longer match meeting times actually yielded smaller effect sizes. Although more research is needed to better understand this finding, it may point to the need for programs to establish realistic expectations around the time commitment to the program (Grossman et al. 2012). Programs that specify day-long activities or multiple hours of relationship-building per session might be over-taxing the commitment of the mentor, youth, or youth's family, raising the risk for relationship dissatisfaction or premature closure. Moreover, expectations for long match meetings might be indicative of more episodic models, through which mentor and youth meetings occur for multiple hours at a time, but over the course of just a few days or weeks, or spaced over large intervals of time. This approach may compromise the capacity of adults and youth to develop the close, supportive one-on-one relationship that is needed to influence youth outcomes, relative to shorter, but more frequent match meetings.

There were no differences observed between school-based and community-based mentoring programs. This is noteworthy, given that school-based mentoring has been the fastest growing program model in recent years (Wheeler et al. 2010). Initially, there was some concern that school-based mentoring relationships may be less influential than relationships forged through community-based programs, since school-based mentoring programs often meet in groups in a single location on school property, and typically have a more limited time commitment (Wheeler et al. 2010). The limited time commitment may, however, protect mentors from burnout and youth from disappointment. These findings are encouraging and indicate that school-based mentoring can be at least as effective as community-based mentoring.

Although these findings present some direction for the establishment of more effective youth mentoring interventions, further research is needed in order to comprehensively and precisely define the types of program structures and settings that are most effective within youth mentoring. It may be the case that the most effective program structure varies depending on factors such as the youth developmental stage or the particular outcome being targeted. For example, in order to maximize the effects of mentoring on youth academic engagement and aggressive behavior at school, a time-limited, school-based mentoring program might be most effective, while a program for addressing youth depression might require a different approach.

## Study Limitations and Strengths

There are a number of limitations to the current analyses. First, meta-analyses are dependent on the availability, type, and quality of evaluations included in the analyses. In the present meta-analysis, only evaluations written in English were coded and analyzed. As a result, although the vast majority of screened studies were written in English, and several studies of mentoring programs in countries outside of North America were included (e.g., Bodin & Leifman; Simões & Alarcão 2014), other rigorous evaluations were excluded purely on the basis of publication language. The present findings therefore may not generalize to mentoring programs in all parts of the world, and future research should explore, for example, the ways in which the role of youth and mentor demographic characteristics might vary across countries. Moreover, although these analyses included a number of methodological factors as moderators of study outcomes, issues such as reporter bias or unreliable or poorly validated measurement tools in the original studies may still have played a role in the observed effect sizes of youth mentoring. Relatedly, moderators could only be tested for studies that reported on these variables, and certain moderators identified as potentially relevant by previous research or theory could not be tested in the current study because of lack of report.

In addition, meta-analyses, by definition, aggregate findings across many populations and program structures, and there was substantial heterogeneity both across studies included in this set of analyses, as well as between effect sizes extracted from the same study. Multiple moderators were tested to attempt to account for this heterogeneity in the current sample of studies; however, further research is needed to more precisely determine which program practices are most effective for which populations of mentors and youth, and for which particular youth outcomes. As this literature grows, future meta-analyses should, when possible, examine more complex interactions between moderating variables—for example, it is possible that mentor and youth demographic characteristics work in concert to influence the impact of the mentoring program (e.g., Blake-Beard et al.

2011; Raposa et al. 2019). Finally, although the present analyses showed no evidence of publication bias, it is important to acknowledge that studies that do not support the effectiveness of mentoring programs might be less likely to appear in peer-reviewed journals, dissertations, or research reports, thereby influencing interpretation of results.

Despite these limitations, this meta-analysis provides the most up-to-date assessment of the impact of youth mentoring, and presents several substantial improvements over previous analyses. It is the first time that these moderators of mentoring program effectiveness have been tested within a meta-analytic sample that includes a comprehensive assemblage of all programs that meet criteria for one-on-one, intergenerational mentoring relationships with youth. Moreover, the present analyses used novel statistical techniques to better account for heterogeneity across effect sizes, as well as several other rigorous methodological approaches (e.g., correction for small samples, weighted effect sizes) that present substantial advantages relative to past meta-analyses, and the analyses included a broader, more scientifically aligned range of outcomes.

## Conclusion

Mentoring programs, which pair youth with caring, non-parental adults with the goal of promoting positive youth development, are an increasingly popular strategy for early intervention with at-risk youth; yet, important questions remain about the extent to which these interventions improve youth outcomes. The present study therefore involved an updated, comprehensive meta-analysis of all mentoring outcome studies to-date, using rigorous inclusion criteria designed to align with developmental theories of youth mentoring. These analyses suggested that youth mentoring programs remain a moderately effective intervention for youth at-risk for a range of psychosocial and academic problems across diverse outcome domains. Moreover, the results of moderation analyses suggest that there may be particular benefits to targeted, time-limited approaches that draw on the service of volunteers or para-professionals with helping experience. Nonetheless, particularly in the context of limited resources, it will be essential to continue to explore for whom mentoring is most effective, and which program practices strengthen (or diminish) the effects of mentoring. Taken together, the current findings provide some support for the efficacy of one-on-one, caring relationships with adults, particularly as a low-cost intervention with the potential to reach large groups of youth and prevent more intensive treatments. Nevertheless, these findings emphasize the need to remain realistic about the potential of mentoring programs as currently implemented, and highlight opportunities for improving the quality and rigor of mentoring practices and evaluation strategies.

**Acknowledgements** This work was supported by funding from the MacArthur Foundation Research Network on Connected Learning and MENTOR: The National Mentoring Partnership.

**Authors' Contributions** E.B.R. led data management and analysis, wrote the first draft of the manuscript, and coordinated draft revisions; J.R. conceptualized and designed the study, oversaw study execution, and contributed to drafts of the manuscript; G.J.J.M.S. and N.C. consulted on study coding and conducted the statistical analyses; J.K. led the development of the outcome coding scheme; S.S. reviewed and contributed to the final draft of the manuscript; and S.B., L.Y.S., S.K., and S.H. conducted literature searches and then identified and coded relevant studies. All authors read and approved the final manuscript.

**Data Sharing and Declaration** The datasets generated and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

## Compliance with Ethical Standards

**Conflict of Interest** The authors declare that they have no conflict of interest.

**Ethical Approval** All research reported on in the manuscript was conducted in compliance with APA ethical principles. The study consisted of secondary analyses of de-identified data, and therefore did not require formal consent or ethics board approval.



## References

Aos, S., Lieb, R., Mayfield, M., Miller, M., & Pennucci, A. (2004). Benefits and costs of prevention and early intervention programs for youth. Olympia, WA: Washington State Institute for Public Policy.

Arnold, M. E., & Cater, M. (2016). From then to now: emerging directions for youth program evaluation. *Journal of Youth Development*, *6*(3), 80–92.

Assink, M., & Wibbelink, C. M. (2016). Fitting three-level meta-analytic models in R: a step-by-step tutorial. *The Quantitative Methods for Psychology*, *12*, 154–174. https://doi.org/10.20982/tqmp.12.3.p154.

Barnet, B., Liu, J., DeVoe, M., Alperovitz-Bichell, K., & Duggan, A. K. (2007). Home visiting for adolescent mothers: effects on parenting, maternal life course, and primary care linkage. *Annals of Family Medicine*, *5*(3), 224–232. https://doi.org/10.1370/afm.629.

Belsky, J., & Cassidy, J. (1994). Attachment and close relationships: An individual-difference perspective. *Psychological inquiry*, *5*(1), 27–30.

Blake-Beard, S., Bayne, M. L., Crosby, F. J., & Muller, C. B. (2011). Matching by race and gender in mentoring relationships: Keeping our eyes on the prize. *Journal of Social issues*, *67*(3), 622–643.

Blakeslee, J. E., & Keller, T. E. (2012). Building the youth mentoring knowledge base: publishing trends and coauthorship networks. *Journal of Community Psychology*, *40*(7), 845–859.

Bogat, G. A., & Liang, B. (2005). Gender in mentoring relationships. In D. L. DuBois & M. J. Karcher (Eds.), *Handbook of Youth Mentoring* (pp. 205–217). Thousand Oaks: Sage Publications.

Bornstein, M. H., Hahn, C. S., & Haynes, O. M. (2010). Social competence, externalizing, and internalizing behavior adjustment from early childhood through early adolescence: developmental cascades. *Development and Psychopathology*, *22*, 717–735.

Bowlby, J. (1988). A secure base: parent-child attachment and healthy human development. Basic Books, USA.

Burns, M. K., Senesac, B. V., & Symington, T. (2003). The effectiveness of the hosts program in improving the reading achievement of children at-risk for reading failure. *Reading Research and Instruction*, *43*(2), 87–103. https://doi.org/10.1080/19388070409558406.

Cavell, T. A., & Elledge, L. C. (2013). Mentoring and prevention science. Handbook of youth mentoring, 29–43.

Cavell, T. A., & Elledge, L. C. (2014). Mentoring and prevention science. In D. DuBois & M. Karcher (Eds.), *Handbook of Youth Mentoring* (2nd Edn, pp. 29–42). Los Angeles: Sage.

Chan, C., Rhodes, J., Howard, W., Schwartz, S., Lowe, S., & Herrera, C. (2013). Pathways of influence in school-based mentoring: the mediating role of parent and teacher relationships. *Journal of School Psychology*, *51*, 129–142. https://doi.org/10.1016/j.jsp.2012.10.001.

Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. *Psychological Methods*, *19*, 211–229.

Cheung, A. C. K., & Slavin, R. E. (2015). How methodological features affect effect sizes in education. *Best Evidence Encyclopedia*. Baltimore, MD: Johns Hopkins University, 1–35.

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

Cummings, L. (2010). Evaluating the influence of participation in a diverse high school-based group mentoring program. Unpublished doctoral dissertation, Georgia State University.

Darling, N., Hamilton, S., Toyokawa, T., & Matsuda, S. (2002). Naturally occurring mentoring in Japan and the United States: social roles and correlates. *American Journal of Community Psychology*, *30*(2), 245–270.

DuBois, D., & Karcher, M. (2013). Youth mentoring in contemporary perspective. In DuBois & M. Karcher (Eds), *The Handbook of Youth Mentoring* (2nd edn). Thousand Oaks: Sage.

DuBois, D. L., Holloway, B. E., Valentine, J. C., & Cooper, H. (2002). Effectiveness of mentoring programs for youth: a meta-analytic review. *American Journal of Community Psychology*, *30*(2), 157–197. https://doi.org/10.1023/A:1014628810714.

DuBois, D. L., Portillo, N., Rhodes, J. E., Silverthorn, N., & Valentine, J. C. (2011). How effective are mentoring programs for youth? A systematic assessment of the evidence. *Psychological Science in the Public Interest*, *12*(2), 57–91. https://doi.org/10.1177/1529100611414806.

Durlak, J. A., & Wells, A. M. (1998). Evaluation of indicated preventive intervention (secondary prevention) mental health programs for children and adolescents. *American journal of community psychology*, *26*(5), 775–802.

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot-based method of using and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463.

Eby, L. T., Allen, T. D., Evans, S. C., Ng, T., & DuBois, D. L. (2008). Does mentoring matter? A multidisciplinary meta-analysis comparing mentored and non-mentored individuals. *Journal of Vocational Behavior*, *72*(2), 254–267. https://doi.org/10.1016/j.jvb.2007.04.005.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315* (7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629.

Everhart, K. (2000). Promoting Resiliency in at-risk children through an integrated adult-child mentoring and character development program: an analysis of mentoring strategies with regard to outcomes. (Unpublished doctoral dissertation). University of South Carolina.

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: unlocking the file drawer. *Science*, *345*, 1502–1505. https://doi.org/10.1126/science.1255484.

Jarjoura, G. R. Tanyu, M, Forbush, J., Herrera, C., & Keller, T. (2018). *Evaluation of the Mentoring Enhancement Demonstration Program: Technical Report*. American Institutes for Research Technical Report. Washington, DC.

Gatto, M. (2006). The effects of a goal-oriented mentoring group on the self-efficacy of children. Unpublished doctoral dissertation, Adelphi University, Garden City, New York.

Garringer, M., Kupersmidt, J., Rhodes, J., Stelter, R., & Tai, T. (2015). *Elements of effective practice for mentoring* (4th edn). Boston: MENTOR: The National Mentoring Partnership.

Grossman, J. B., Chan, C. S., Schwartz, S. E. O., & Rhodes, J. E. (2012). The test of time in school-based mentoring: the role of relationship duration and re-matching on academic outcomes. *American Journal of Community Psychology*, *49*(1-2), 43–53. https://doi.org/10.1007/s10464-011-9435-0.

Grossman, J. B., & Rhodes, J. E. (2002). The test of time: predictors and effects of duration in youth mentoring relationships. *American Journal of Community Psychology*, *30*(2), 199–219. https://doi.org/10.1080/13611267.2013.855864.

Hanlon, T. E., Simon, B. D., O'Grady, K. E., Carswell, S. B., & Callaman, J. M. (2009). The effectiveness of an after-school program targeting urban African American youth. *Education and Urban Society*, *42*(1), 96–118. https://doi.org/10.1177/0013124509343144.

Hedges, L., & Olkin, I. (1985). *Statistical models for meta-analysis*. New York: Academic Press.

*Herrera, C., DuBois, D. L., & Grossman, J. B. (2013). The role of risk: mentoring experiences and outcomes for youth with varying risk profiles. New York, NY: MDRC.

*Herrera, C., Grossman, J. B., Kauh, T. J., & McMaken, J. (2011). mentoring in schools: an impact study of big brothers big sisters school-based mentoring. *Child Development*, *82*(1), 346–361.

*Herrera, C. Kauh, T. J., Cooney, S. M., Grossman, J. B., & McMaken, J. (2008). High school students as mentors: findings from the big brothers big sisters school-based mentoring impact study. Report.

Herrera, C., Sipe, C. L., & McClanahan, W. S. (2000). *Mentoring school-age children: Relationship development in community-based and school-based programs*. Philadelphia: Public/Private Ventures.

Jent, J. F., & Niec, L. N. (2009). Cognitive behavioral principles within group mentoring: a randomized pilot study. *Child and Family Behavioral Therapy*, *31*(3), 203–219. https://doi.org/10.1080/07317100903099258.

Jolliffe, D., & Farrington, D. P. (2007). Examining the relationship between low empathy and self-reported offending. *Legal and Criminological Psychology*, *12*(2), 265–286.

Kanchewa, S. S., Yoviene, L. A., Schwartz, S. E. O., Herrera, C., & Rhodes, J. E. (2016). Relational experiences in school-based mentoring: the mediating role of rejection sensitivity. Youth & Society, 1–22. https://doi.org/10.1177/0044118X16653534

*Karcher, M. J., Davis, C., Powell, B. (2002). The effects of developmental mentoring on connectedness and academic achievement. *The School Community Journal*, *12*(2), 35–50.

Karcher, M. J., Nakkula, M. J., & Harris, J. (2005). Developmental mentoring match characteristics: correspondence between mentors' and mentees' assessments of relationship quality. *Journal of Primary Prevention*, *26*(2), 93–110. https://doi.org/10.1007/s10935-005-1847-x.

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, *22* (17), 2693–2710.

Kupersmidt, J. B., Stump, K. N., Stelter, R. L., & Rhodes, J. E. (2017a). Predictors of premature match closure in youth mentoring relationships. *American Journal of Community Psychology*, *59*(1-2), 25–35.

Kupersmidt, J. B., Stelter, R. L., Rhodes, J. E., & Stump, K. N. (2017b). Enhancing mentor efficacy and preparedness through web-based pre-match training. *The Journal of Nonprofit Education and Leadership*, *7*(3), 197–216. https://doi.org/10.18666/JNEL-2017-V7-I3-794.

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

Markus, H., & Nurius, P. (1986). Possible selves. *American psychologist*, *41*(9), 954–969.

Morrow-Howell, N., Jonson-Reid, M., McCrary, S., Lee, Y., & Spitznagel, E. (2009). *Evaluation of Experience Corps: Student Reading Outcomes*. St. Louis, Missouri: Center for Social Development: Washington University.

National Collaborating Centre for Methods and Tools (2008). *Quality assessment tool for quantitative studies*. Hamilton: McMaster University. http://www.nccmt.ca/knowledge-repositories/search/14.

Parra, G. R., DuBois, D. L., Neville, H. A., Pugh-Lilly, A. O., & Povinelli, N. (2002). Mentoring relationships for youth: investigation of a process-oriented model. *Journal of Community Psychology*, *30*(4), 367–388.

Raposa, E. B., Ben-Eliyahu, A., Olsho, L. E., & Rhodes, J. (2019). Birds of a feather: is matching based on shared interests and characteristics associated with longer youth mentoring relationships? *Journal of Community Psychology*. 1–13.

Raposa, E. B., Dietz, N., & Rhodes, J. E. (2017). Trends in Volunteer mentoring in the United States: analysis of a decade of Census Survey Data. *American Journal of Community Psychology*, *1*, 1–12. https://doi.org/10.1002/ajcp.12117.

Raposa, E. B., Rhodes, J. E., & Herrera, C. (2016). The impact of youth risk on mentoring relationship quality: do mentor characteristics matter? *American Journal of Community Psychology*, *57*, 320–329. https://doi.org/10.1002/ajcp.12057.

Rhodes, J. E. (2005). A theoretical model of youth mentoring. In D. L. DuBois & M. J. Karcher (Eds.), *Handbook of youth mentoring* (pp. 30–43). Thousand Oakes: Sage Press.

Rhodes, J. E., Lowe, S., Litchfield, L., & Samp-Walsh, K. W. (2008). The role of gender in youth mentoring relationship satisfaction and duration. *Journal of Vocational Behavior*, *72*(2), 183–192. https://doi.org/10.1016/j.jvb.2007.09.005.

Rhodes, J. E., Reddy, R., & Grossman, J. B. (2005). The protective influence of mentoring on adolescents' substance use: direct and indirect pathways. *Applied Developmental Science*, *9*(1), 31–47.

Rhodes, J. E., Reddy, R., Grossman, J. B., & Lee, J. M. (2002). Volunteer mentoring relationships with minority youth: an analysis of same- versus cross-race matches. *Journal of Applied Social Psychology*, *32*(10), 2114–2133. https://doi.org/10.1111/j.1559-1816.2002.tb02066.x.

Rogoff, B. (1978). Spot observation: an introduction and examination. *Quarterly Newsletter of the Institute for Comparative Human Development*, *2*(2C), 21–26.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*, 639–641.

Ruzek, E. A., Hafen, C. A., Allen, J. P., Gregory, A., Mikami, A. Y., & Pianta, R. C. (2016). How teacher emotional support motivates students: the mediating roles of perceived peer relatedness, autonomy support, and competence. *Learning and Instruction*, *42*, 95–103.

Sánchez, B., Reyes, O., & Singh, J. (2006). Makin'it in college: the value of significant individuals in the lives of Mexican American adolescents. *Journal of Hispanic Higher Education*, *5*(1), 48–67.

Simões, F., & Alarcão, M. (2014). Promoting well-being in school-based mentoring through basic psychological needs support: Does it really count? *Journal of Happiness Studies*, *15*(2), 407–424.

Splendelow, J. S., Simonds, L. M., & Avery, R. E. (2017). The relationship between co-rumination and internalizing problems: a systematic review and meta-analysis. *Clinical Psychology & Psychotherapy*, *24*(2), 512–527.

Spilt, J. L., Hughes, J. N., Wu, J. Y., & Kwok, O. M. (2012). Dynamics of teacher-student relationships: Stability and change across elementary school and the influence on children as academic success. *Child development*, *83*(4), 1180–1195.

Sutton, A. J., Duval, S., Tweedie, R., Abrams, K. R., & Jones, D. R. (2000). Empirical assessment of effect of publication bias on meta-analyses. *BMJ*, *320*, 1575–1577.

Tabachnik, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics (*6th edn). Boston: Allyn and Bacon.

Tanner-Smith, E. E., Durlak, J. A., & Marx, R. A. (2018). Empirically based mean effect size distributions for universal prevention programs targeting school-aged youth: a review of meta-analyses. *Prevention Science*, *19*(8), 1091–1101.

*Taussig, H. N., & Culhane, S. E. (2010). Impact of a mentoring and skills group program on mental health outcomes for maltreated children in foster care. *Archives of Pediatric & Adolescent Medicine*, *164*(8), 739–746.

Thompson, R. B., Corsello, M., McReynolds, S., & Conklin-Powers, B. (2013). A longitudinal study of family socioeconomic status (SES) variables as predictors of socio-emotional resilience among mentored youth. *Mentoring & Tutoring: Partnership in Learning*, *21*(4), 378–391. https://doi.org/10.1080/13611267.2013.855864.

Tolan, P., Henry, D., Schoeny, M., & Bass, A. (2008). *Mentoring interventions to affect juvenile delinquency and associated problems*. Chicago: The University of Illinois at Chicago.

Van Dam, L., Smit, D., Wildschut, B., Branje, S. J. T., Rhodes, J. E., Assink, M., & Stams, G. J. (2017). Does natural mentoring matter: a multilevel meta-analysis on the association between natural mentoring and youth outcomes. American Journal of Community Psychology, *62*, 203–220.

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2014). Meta-analysis of multiple outcomes: a multilevel approach. *Behavior Research Methods*, *47*, 1274–1294. https://doi.org/10.3758/s13428-014-0527-2.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 1–48.

Wheeler, M. E., Keller, T. E., & DuBois, D. L. (2010). Review of three recent randomized trials of school-based mentoring: making sense of mixed findings. *Social Policy Report*, *24*(3), 3–21.

Woods, C., & Preciado, M. (2016). Student-mentor relationships and students' college attitudes. *Journal of EduWhecation for Students Placed at Risk (JESPAR)*, *21*(2), 90–103. https://doi.org/10.1080/10824669.2015.1127767.

**Elizabeth B. Raposa** is an Assistant Professor in the Department of Psychological Sciences at William and Mary. Her research investigates how early life stressors influence risk for mental health problems across development, and how close relationships with parents, peers, or others mitigate the negative impact of stressors on youth.

**Jean Rhodes** is the Frank L. Boyden Professor of Psychology and the Director of the Center for Evidence-Based Mentoring at the University of Massachusetts Boston. She has devoted her career to understanding the role of both natural and assigned mentoring relationships in the lives of disadvantaged youth, including youth in the foster care system.

Geert-Jan Stams is a Professor of Forensic Child and Youth Care Sciences at the University of Amsterdam. His major research interests include meta-analysis and youth delinquency.

**Noel Card** is a Professor in Human Development and Family Studies at the University of Connecticut. His major research interests include meta-analysis, positive psychology, and youth development.

**Samantha Burton** is a doctoral student in clinical psychology at the University of Massachusetts, Boston. Her major research interests include youth mentoring, social capital, and implicit bias.

**Sarah Schwartz** is an Assistant Professor of Psychology at Suffolk University. Her major research interests include mentoring, social capital, and trauma.

**Laura A. Yoviene Sykes** is a postdoctoral fellow at Yale University School of Medicine. Her major research interests include mentoring, attachment theory, and young adult psychopathology.

**Stella Kanchewa** is a research and teaching psychologist at the University of Kentucky, Louisville. Her major research interests include youth mentoring and assessment.

**Janis Kupersmidt** is a Senior Research Scientist at iRT. Her major research interests include mentoring, peer relationships, and prevention science.

**Saida Hussain** is a Research Scientist at the United States Government Accountability Office. Her major research interests include youth mentoring, positive youth development, and social capital.