Received: 10 January 2021 | Revised: 13 October 2021 | Accepted: 26 October 2021

DOI: 10.1002/ajcp.12579

EMPIRICAL REVIEW

Cross-age peer mentoring for youth: A meta-analysis

Samantha Burton[1] | **Elizabeth B. Raposa**[2] | **Cyanea Y. S. Poon**[1] | **Geert Jan J. M. Stams**[3] | **Jean Rhodes**[1]

[1]Department of Psychology, University of Massachusetts, Boston, Massachusetts, USA

[2]Department of Psychology, Fordham University, Bronx, New York, USA

[3]Social and Behavioural Sciences, University of Amsterdam, Amsterdam, The Netherlands

Correspondence

Elizabeth B. Raposa, Department of Psychology, Fordham University, 441 E. Fordham Rd, Bronx, NY 10458, USA.
Email: eraposa@fordham.edu

Abstract

Although most mentoring programs for youth are structured around intergenerational relationships, a growing number of programs rely on cross-age peer mentoring. Such programs capitalize on the availability of youth mentors to promote positive outcomes in younger peers. This study used a multilevel meta-analytic approach to estimate the effect size of cross-age peer mentoring programs and evaluate potential moderators of peer mentoring program effectiveness. Analyses included six studies and revealed a medium-sized overall effect of cross-age peer mentoring programs ($g = 0.45$). Several characteristics moderated effect sizes, with larger effects for programs that were conducted outside of the school setting (i.e., weekend, summer, or in community settings), conducted in urban settings, and had moderate/high levels of adult oversight and supervision. Results highlight the potential benefits of cross-age peer mentoring for youth.

KEYWORDS

cross-age, meta-analysis, peer mentoring, youth outcomes

Highlights

- Analyses revealed a medium-sized overall effect of cross-age peer mentoring programs ($g = 0.45$).
- Cross-age peer mentoring can be an accessible intervention for promoting positive youth outcomes.
- Peer mentoring programs with moderate to high levels of adult oversight and supervision had larger effects.
- This study is the first meta-analytic assessment of the impact of cross-age peer mentoring programs.

THEORIES OF PEER INFLUENCE

Cross-age peer mentoring programs are built on the assumption that same-generation peers are influential for youth social and cultural development. These peer influences become evident by early childhood, when the amount of time peers spend together begins to exceed the amount of time youth spend with their parents (Ellis et al., 1981; Santrock, 2019). Harris' (1995) "group socialization theory" suggests that children identify with a peer group and adapt their own behavior to fit in with the behavioral norms of the group, while distinguishing themselves from other groups. Although adults may possess authority based on age and position, older youth can provide younger youth with status and social connection (Smith, 2011), as well as guidance on how to speak, act, and dress within their peer group (Harris, 1998). Consequently, older peers are thought to be uniquely situated to provide a relationship within which the mentee can value, respect, and idealize the mentor.

The argument for cross-age peer mentors is further strengthened by theoretical models of the normative process of establishing independence from adult authority figures during adolescence (Miller-Johnson & Constanzo, 2004). During middle childhood and early adolescence, many youth naturally begin to chafe against conventional adult norms of behavior (Hirschi, 1969; Jessor & Jessor, 1977). As such, advice and guidance from adults is less likely to be assimilated into the youth's intrinsic sense of self-identity. In fact, when youth structure their own behaviors, they often do so in opposition to these adult conventions, striving for unconventionality (Karcher, 2005a). In contrast, when

older peers model and reward these conventions, they may not be met with the same resistance; youth can adopt these conventions as part of their self-concept while establishing independence from adults (Karcher, 2005a). Thus, cross-age peer mentoring may offer a more efficient and impactful opportunity to reward prosocial attitudes and behaviors through older peer modeling and support.

Despite the potential for positive influences of peer-based mentoring, there are also risks. Some argue that adolescent mentors may not be sufficiently mature to mentor their younger peers, and may not have the skills to scaffold the mentee's emotional development while practicing higher-order cognitive and perspective-taking skills (Adler, 1964; Selman, 1980). Another concerning phenomena regarding peer influence is "peer contagion," the process by which peers exert mutual influence on each others' negative behaviors (e.g., delinquency, aggression, and substance use) in a way that undermines emotional and behavioral development (Dishion & Tipsord, 2011; Stevens & Prinstein, 2005). Peer contagion occurs when groups of older peers model deviant attitudes or behaviors (Dishion & Tipsord, 2011; O'Donnell & Williams, 2013), and can lead to iatrogenic effects of education and intervention programs that bring together high-risk youth (e.g., crime prevention programs, treatment for antisocial youth; Feldman, 1992; McCord, 2003). Although meta-analyses of group interventions for youth reveal overall positive effects, and well-supervised interventions do not seem to produce negative effects overall (Dishion & Tipsord, 2011; M. Lipsey, 2006), more research is needed that investigates the conditions under which peer contagion may occur and contribute to more negative youth outcomes (O'Donnell & Williams, 2013). This is particularly true in peer mentoring programs, which often provide opportunities for mentors and mentees to interact with one another in group settings.

CROSS-AGE PEER MENTORING PROGRAMS

Consistent with the idea that peer relationships can have varied impacts on youth, initial findings from cross-age peer mentoring programs point to mixed effects. For example, evaluations of the Cross-age Mentoring Program (CAMP), which incorporates academic instruction as part of the intervention, showed mentee improvement in school connectedness and spelling achievement (Karcher, 2008; Karcher et al., 2002), but no changes in grades or other achievement outcomes (Karcher, 2008). Likewise, in a large-scale evaluation of Big Brother Big Sisters (BBBS), researchers found that cross-age peer mentored youth demonstrated gains in social acceptance, parent relationship quality, and assertiveness, relative to adult-mentored youth, but found that mentees with older peer mentors benefitted less or not at all in areas of misconduct, classroom effort, grade point average (GPA), and intentions to go to college (Herrera et al., 2008). In a comprehensive review of cross-age peer mentoring programs, Karcher and Berger (2017) concluded that there was some evidence of benefits to youth who participate in cross-age peer mentoring programs, particularly with respect to social support, self-esteem, and school connectedness. However, they also cautioned that findings were mixed, with fewer studies showing an effect on academic and behavioral outcomes. They noted that studies of peer mentoring tend to vary considerably in design and quality.

POTENTIAL MODERATORS OF PEER MENTORING EFFECTIVENESS

Thus, important questions remain about the overall effectiveness of cross-age peer mentoring and whether a wide range of youth, mentor, match, program, and methodological characteristics may moderate program effects. Research on intergenerational mentoring relationships has identified several youth characteristics that may be important predictors of program effectiveness for cross-age peer mentoring. For example, some youth demographic characteristics, such as youth gender composition of the program, appear to play a role in predicting effect sizes in intergenerational mentoring programs (DuBois et al., 2002, 2011; Raposa et al., 2019). In addition, some studies have demonstrated that mentees with greater behavioral difficulties, like conduct problems or poor academic performance, tend to have lower-quality mentoring relationships, more inconsistent match meetings, and a higher likelihood of experiencing an early match closure due to their mentor dropping out (Herrera et al., 2008; Karcher, 2005b; Karcher & Lindwall, 2003). Yet, other studies have shown *stronger* outcomes for programs serving youth with greater levels of individual and environmental risk (DuBois et al., 2011), and for youth who initially report higher risk for negative outcomes (DuBois et al., 2002; Poon, 2019), suggesting that further research is needed on the impact of youth baseline risk factors on mentoring relationships. Moreover, none of these studies have included specific tests of the impact of youth characteristics on cross-age peer mentoring relationships.

A range of mentor characteristics have also been linked to the effectiveness of mentoring interventions in general, although these moderators have rarely been tested within the context of cross-age peer mentoring. For example, research has linked mentors' prosocial attitudes to better mentoring program outcomes (Karcher & Lindwall, 2003). In addition, a recent meta-analysis of intergenerational youth mentoring highlighted the potential impact of mentors' demographic characteristics by demonstrating larger effects in programs that had a higher percentage of male mentors (Raposa et al., 2019). In one study involving cross-age peer mentoring, teen mentors with more positive attitudes toward youth appear to be more effective with academically disconnected mentees than mentors with less positive attitudes toward youth (Herrera et al., 2008).

There is also substantial diversity in program practices across cross-age peer mentoring programs, and these program practices are likely to have important effects on intervention outcomes. For example, one potential moderator of the impact of peer mentoring programs involves adult oversight and involvement in the intervention. Past studies suggest that level of staff support is positively associated with peer mentors' views of relationship quality and program satisfaction (Herrera et al., 2008), and that increased time spent in training and higher-quality mentor training can also positively impact outcomes (Herrera et al., 2008; Karcher et al., 2011). Furthermore, programs that promote mentees' parents' involvement through family events (Karcher et al., 2005), and that incorporate both individual and larger group mentoring activities (Herrera et al., 2008), demonstrate greater peer mentor satisfaction and longer match duration. Together these findings indicate that level of adult support can influence youth mentoring effectiveness.

Finally, the methodological approach of the study is an important predictor of effect sizes in meta-analyses across fields. Specifically, research shows that studies employing random assignment tend to yield smaller effect sizes than those employing less rigorous quasi-experimental designs (A. Cheung & Slavin, 2015). Additionally, published studies tend to report greater effect sizes than unpublished reports due to biases in publishing significant results (A. Cheung & Slavin, 2015). These potential publication biases are important to consider when conducting a thorough meta-analysis.

CURRENT STUDY

Although several meta-analyses of intergenerational mentoring programs have been conducted (DuBois et al., 2002, 2011; Raposa et al., 2019), there have been no meta-analyses of cross-age peer mentoring. To address gaps in the existing literature, the current meta-analysis examined the impact of cross-age peer mentoring using all relevant outcome studies of cross-age, one-on-one peer mentoring programs for youth that were written in English. Consistent with the Karcher and Berger (2017) review, stringent inclusion criteria ensured that analyses examined cross-age peer mentoring programs that were designed to improve youth outcomes through a supportive relationship in which there was a difference of two or more years in age between an older mentor and a younger mentee of the same generation. Using a multilevel meta-analytic approach, the analyses (1) estimated the overall effect size of cross-age peer mentoring programs, as well as within- and between-study variability in effect sizes; (2) tested whether the effects of cross-age peer mentoring were different across diverse outcome categories (e.g., school-related vs. psychological outcomes); (3) examined whether the size of program effects were moderated by key youth characteristics, mentor characteristics, program characteristics, and methodological characteristics; and (4) tested the role of publication bias in the calculated overall effect size.

METHODS

Study selection

A comprehensive search of the literature published before April 2019 was conducted to identify evaluations of cross-age peer mentoring programs (see Figure 1). 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: Peer mentoring, Cross-age peer mentoring, Peer mentoring + evaluation, Peer mentoring + intervention, Peer mentoring + outcomes, Peer mentoring + effects, Peer mentoring + randomized control trial, Peer mentoring + experimental. Moreover, prior cross-age one-on-one peer mentoring reviews and intergenerational mentoring meta-analyses were manually searched to identify additional articles.

Duplicate studies were screened out before evaluation for inclusion. To be considered for inclusion in the final sample, studies were required to meet the following criteria: (1) A formal mentoring program, with mentoring defined as an older youth (at least two years older) acting in a nonprofessional helping capacity with a specific younger person to promote positive youth outcomes through the relationship. (2) An evaluation with a comparison group, including randomized controlled trials and quasi-experimental studies.

Studies were excluded from the meta-analysis if they met any of the following exclusion criteria: (1) studies with adult mentors (18 years of age or older) or that combined adult and cross-age peer mentors without a separate analysis of each; (2) mentees and mentors who had less than a two-year age difference (the required two-year age gap between mentor and mentee was based on Karcher and Berger's 2017 definition of cross-age youth mentoring, which emphasizes the importance of older youth mentors in fulfilling mentoring roles similar to those in adult-youth mentoring that same-age peer relationships might not, such as acting as a role model, providing support, and offering guidance to their mentee); (3) only group mentoring; (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 did not respond to requests for

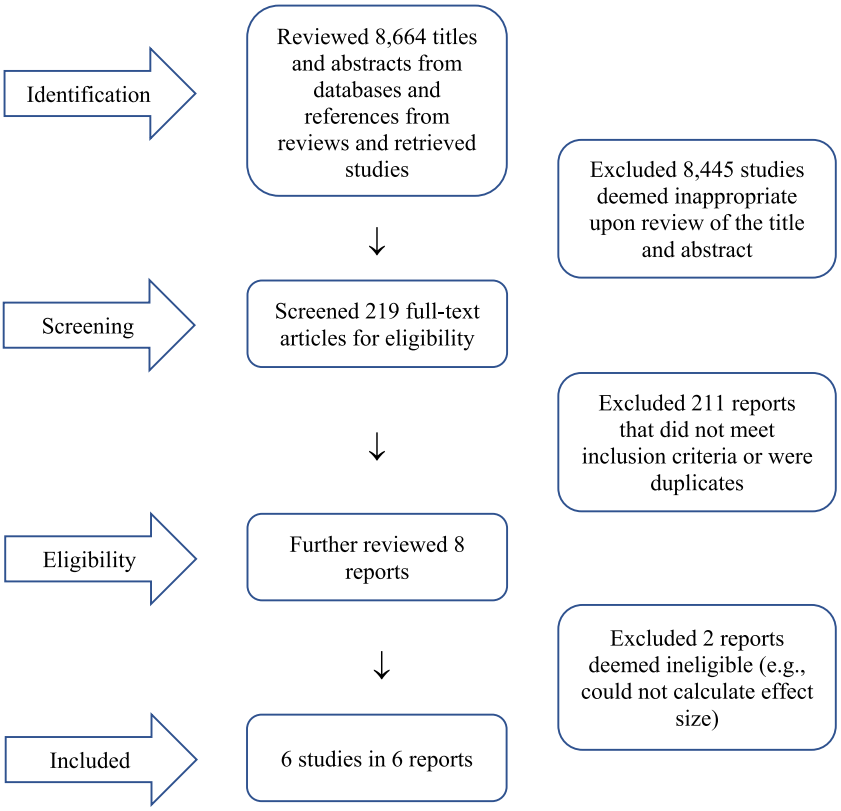

FIGURE 1 Flowchart of the study selection process. The comprehensive search for studies utilized computerized database searches (PsycINFO, ERIC, and ProQuest), as well as a manual search of other resources. Studies were screened for inclusion and exclusion criteria, and authors were contacted if additional data were 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 six studies for analysis [Color figure can be viewed at wileyonlinelibrary.com]

additional information within one month of the request; and (8) the study was written in a language other than English.

This procedure yielded six studies for analysis (see Table 1 for a description of all included studies). Although this is a relatively small sample of studies, the minimum number of studies for a meta-analysis is only three (Treadwell et al., 2006), and most published meta-analyses include less than nine studies (Lau et al., 2006). In addition, meta-analytic evaluations of mentoring programs have been conducted for small samples in the past (e.g., three studies were included in Wheeler et al., 2010), demonstrating that this analytic method can be useful and informative even with a limited number of studies included.

STUDY CODING PROCEDURES

Outcomes

Studies were coded for mentee, mentor, match, program, and methodological characteristics by two raters. Raters adhered to a coding manual, which outlined coding procedures and codes for each characteristic. Raters with advanced statistical experience coded study outcome data and calculated effect sizes. The lead coders attended a training led by a researcher with expertize 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 the raters discussed and resolved difficulties or discrepancies in coding and effect size calculation. All studies were double-coded and discrepancies in coding were resolved through joint review of study details and consultation of past literature.

Outcomes for each study were noted and coded for several characteristics, such as information source and statistical details (e.g., sample size, means, and standard deviations). These coded outcome characteristics were utilized to calculate an effect size for each outcome. Outcomes were coded as belonging to one of the following five broad categories: psychological, social, cognitive, health, or school. Additionally, each broad outcome was coded according to a more specific set of sub-categories, termed "narrow outcomes" (i.e., externalizing symptoms, internalizing symptoms, self-regulation problems, overall mental health, substance use, physical health, general well-being, executive functioning, self-cognition, academic functioning, school engagement, extracurricular engagement, social skills, social support, peer relationship, teacher relationship, and parent relationship). These categories have been used in past meta-analyses of youth mentoring studies (e.g., Christensen et al., 2020; Raposa et al., 2019) and were derived from review of the developmental psychopathology and positive youth development literatures by three Ph.D.-level researchers with expertize in the areas of youth development and clinical psychology. This team of researchers completed an iterative process of (1) collaborating on a review of the relevant literature; (2) independently deriving codes that seemed to best capture the

TABLE 1 Studies included in the current meta-analysis

Name	Description	Findings
1. Big Brothers Big Sisters of America, High School Bigs study (Herrera et al., 2008; data taken from What Works Clearinghouse Review, 2013)	Relationship-focused school-based mentoring programs with high school mentors supported by local BBBS agencies. Volunteer mentors met with students at their school for approximately 1 hour per week during or after school, with a general focus on social and academic activities.	Youth matched with a high school mentor did not show significant improvement on a variety of school-related self- and teacher-report measures (e.g., misconduct, classroom effort, etc.) when compared with non-mentored youth. Several moderators contributed to increased benefits of having a teenage mentor, namely mentor training (higher amount and better quality) and staff support (better perceived quality and increased frequency of communication).
2. Cross-Age Mentoring Program (CAMP) Cross-Campus Model (Karcher, 2005b)	Developmental mentoring program targeting social and school connectedness. Meetings included a whole-group icebreaker activity, one-to-one informal conversation and discussion time, a structured dyadic activity from a connectedness curriculum, and short unstructured time to interact in the larger group.	Findings indicated that mentored youth reported higher scores on connectedness to school and parents at post-test than the control group. Mentor attendance, but not mentee attendance, was positively associated with pre-to-post changes in mentees' self-reported rule compliance, social skills, and self-esteem, suggesting exposure to the curriculum (i.e., mentee attendance) was less predictive of program changes than was the mentor's presence.
3. Cross-Age Mentoring Program (CAMP) Outreach Model (Karcher et al., 2002)	Developmental mentoring program targeting social and school connectedness. Monthly Saturday events for a full academic year plus a summer enrichment program. Meetings were structured by a curriculum for the mentor-mentee relationships, academic skills development activities, and connectedness activities, in addition to unstructured time to interact in the larger group.	At 1 year (posttest) the mentored youth reported higher scores on connectedness to parents and spelling achievement. Analyses revealed that improvements in spelling achievement were fully explained by gains in connectedness to parents, suggesting that academic benefits from program participation were largely due to gains in connectedness to parents that resulted from program participation.
4. Children Teaching Children (CTC) Program (Sheehan et al., 1999)	Mentoring program focused on violence prevention among youth living in a violent neighborhood. Program goals were to develop a cross-age mentoring relationship, structured by violence prevention activities to modify violence attitudes and behaviors among preadolescents. The teenage mentors in the program designed and presented lessons to teach younger children about violence prevention.	On the first measure assessing "exposure to violence and/or acceptance of violence," at the end of the study, the intervention group reported lower scores than control subjects. On a second measure of acceptance of violence, differences between the intervention and comparison group emerged at both 9 and 18 months, favoring mentees. Teacher ratings of youth behavior showed significantly worse ratings for control group subjects compared to the intervention group, post-treatment.
5. Just for Kids! Mentoring to promote healthy diet and physical activity among children in Appalachia (Smith, 2011)	Mentoring program focused on promotion of healthier patterns of dietary intake and physical activity in a rural population with high rates of childhood obesity. Children received curriculum delivered by trained high school-age teen mentors. Curriculum addressed the roles of exercise, daily activity, and food in promoting health and encouraged children to set reasonable behavioral goals. Also addressed self-acceptance, processing emotions, assertiveness, and positive self-evaluation.	Results indicate that aspects of the Just for Kids! curriculum were effective in impacting children's dietary intentions and BMI percentiles. The curriculum did not impact intention to engage in regular physical activity. The greatest gain for the intervention group was in self-efficacy toward physical activity. Overall, a change in attitudes toward eating healthfully trended toward improvement for less than half of the children in the intervention group. Gains were noted in perceived support from others such as family members to engage in physical activity and eat healthfully among children in the teen-mentored group.
6. "Will and Skill" A mentor program for improving the academic attainment of Black adolescent males (Tomlin, 1994)	Mentoring program targeting improved academic attainment of black adolescent males with high academic risk ratings.	Mentored youth showed significantly better posttest results on measures of self-efficacy, grade point average, and teacher conduct

(Continues)

TABLE 1 (Continued)

Name	Description	Findings
	Mentoring focused on the establishment of a "skill" domain to help students acquire strategies for self-regulatory skills and academic success. Format of mentor/mentee meetings included discussions about mentees' school performance and progress reports, as well as teaching self-regulatory learning strategies. Mentors verbally shared how they use strategies, modeled the strategies, and asked mentees about their strategy usage.	ratings compared to wait list control youth. No significant differences were found for measures of unexcused absences, office referrals, suspensions, and self-perception for scholastic competence, social acceptance, behavioral conduct, and global self-worth.

set of outcomes coded for this meta-analysis; and (3) meeting several times to derive a two-tiered set of consensus codes that could be applied. This tiered coding system 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 the outcome type, the following characteristics were coded as potential moderators of program effect sizes.

Mentee characteristics

Mentee gender, age, grade, race/ethnicity (Hispanic/Latino, Black/African American, White, Asian, Native Hawaiian/Other Pacific Islander, American Indian/Alaska Native, Multiracial, and "other"), and mentee sample size were coded and examined as potential moderators of program effectiveness. Several variables were also coded as indicators of youth risk. First, as a proxy for low socioeconomic status (SES), the percent of mentees receiving free or reduced-price lunch was noted for each study. Other coded indicators of risk included 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). In addition, coders rated whether each study was designed for one of the following specific populations of youth: general (un-selected) population, racial/ethnic minority youth, youth from single-parent households, youth from low-SES families, foster care youth, or youth with multiple risk indicators. Finally, raters coded whether mentees received an incentive for their participation in the mentoring program (e.g., course credit and payment).

Mentor characteristics

Mentor gender, age, grade, and race/ethnicity (Hispanic/Latino, Black/African American, White, Asian, Native Hawaiian/Other Pacific Islander, American Indian/Alaska Native, Multiracial, and "other") were examined as predictors of mentoring program effectiveness. In addition, raters coded whether mentors had previous mentoring experience, whether mentors received an incentive for their participation in the mentoring program (e.g., course credit and payment), and whether participation in the mentoring program was voluntary.

Match characteristics

Raters coded match characteristics including percentages of cross-race and cross-sex matches; whether mentoring dyads were intentionally matched based on sex, race, or interests; whether mentors and youth came from the same communities; and average age difference between mentors and mentees.

Program characteristics

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 expected overall program length). Raters also coded for level of adult oversight and supervision (i.e., low and moderate-high) that mentors received throughout their mentoring relationship. This code involved a judgment based on the amount of supervision by staff and parents of mentoring activities (e.g., staff/parents present during none, some, or all mentoring activities), as well as whether the program offered structured intervention support (e.g., staff-mentor meetings to debrief mentoring interventions). In addition, the primary goal of the mentoring program was coded for each study: nonspecific/general positive youth development, improving academic performance, reducing behavioral problems, reducing psychosocial problems, or improving health.

Raters also coded whether mentoring was the sole intervention of the program, or whether it was the primary intervention in the context of other program components. Raters coded whether there was a family component to

each program (e.g., family events/activities), and the program's geographical location (i.e., urban, suburban, rural, or mixed), as well as the primary program site (i.e., school [during or after school day in school setting] vs. other [i.e., weekend, summer, or community-based]). Finally, raters coded the type of mentoring intervention in terms of whether the intervention was solely conducted as a one-to-one relationship or whether the program incorporated both one-to-one and group mentoring activities.

Methodological characteristics

Several aspects of each study's research design were coded 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 was noted. In addition, each study's design was coded as a randomized controlled trial or a quasi-experimental design. The control group for each study was coded as "no treatment" if the control group did not receive any intervention (e.g., a waitlist control), and "treatment as usual" if the control group received other services offered by a program (e.g., tutoring and social services). The source of outcome information (i.e., youth, parent, school record, teacher, or other reporter) was also coded. Finally, a structured rating of study quality (i.e., weak, moderate, or strong) was assigned to each study 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

When means, standard deviations, sample sizes, or other information necessary for the calculations were not reported, study authors were contacted for additional information. 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* to adjust for differences in sample size (Hedges & Olkin, 1985).

Given that more than one effect size was calculated for each study, a three-level approach to meta-analysis was applied 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 (M. W. L. 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 & 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, participant, 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 metaphor-package, using a multilevel random effects model (Assink & 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 & Wibbelink, 2016). Each continuous moderator was centered around its mean, and dichotomous dummy variables were created for all categorical variables (Tabachnik & 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 and sensitivity analyses

There is still no valid and reliable way to examine publication bias in multilevel meta-analysis; therefore, the present analyses tested for publication bias in several ways. First, analyses examined differences in effect sizes between dissertations, program reports, and published journal articles. Second, both a funnel plot and a trim-and-fill analysis (Duval & Tweedie, 2000), were conducted with the function "trimfill" in the metaphor-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 nonsignificant 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.

Finally, given the small sample size of studies, sensitivity analyses were conducted to investigate the robustness of the overall results. The effect sizes were recalculated six times, each time removing a different study, to examine the influence of each individual study on the overall effect size (Viechtbauer & Cheung, 2010).

RESULTS

Average effect of cross-age peer mentoring for youth

There were six studies providing estimates of effect sizes of the impact of cross-age peer mentoring, with a combined sample size of 685 mentees. Characteristics of each study are presented in Table 2 and descriptive information for coded moderators is presented in Table 3. The average effect size across all 6 studies and all outcomes was $g = 0.45$ ($p = .003$; 95% CI: 0.16 to 0.74). This is a statistically significant medium effect size by Cohen's (1988) guidelines. Analyses revealed that there was significant heterogeneity across studies ($\sigma^2_{\text{Level 3}} = .11$, $p < .0001$), but no significant variability between effect sizes extracted from the same study ($\sigma^2_{\text{Level 2}} = .00$, $p < .0001$). Notably, 83% of the variance among effect sizes was accounted for by the between-study level, while random sampling error accounted for 17% of the variance.

DIFFERENCES IN EFFECTS BASED ON YOUTH OUTCOME TYPE

Analysis of differences in outcome type that might have accounted for heterogeneity across effect sizes are presented in Table 4. Results showed no significant differences in effect sizes across the five broad outcome categories: school, social, health, cognitive, and psychological outcomes ($F[1, 41] = .38$, $p = .54$). Regarding the narrow outcome sub-types, only the subtypes that were coded (i.e., relevant for the outcomes in the included studies) were included in the analyses. Results revealed no substantial variability across the more precise, narrow outcome subtypes ($F[1, 41] = 2.37$, $p = .12$). That is, there was no significant variability among the psychological outcomes (externalizing symptoms and overall mental health), health outcomes (substance use and physical health), cognitive outcomes (self-cognition), school outcomes (academic functioning and school engagement), or social outcomes (social skills and overall social support).

MODERATORS OF MENTORING EFFECTIVENESS

Results of moderator analyses on between-study youth, mentor, program, and study/methodological characteristics are summarized in Table 4. Given the low sample

TABLE 2 Study characteristics

Reference	*N*	# ES	Mean ES	Study type	Study design	Length (months)	Study quality	Primary focus	Broad outcomes
Herrera et al. (2008)	416	19	0.03	Report	RCT	9	Moderate	General	School, social, psychological, health
Karcher et al. (2002)	26	4	0.82	Journal article	RCT	12	Moderate	Psychosocial	School, cognitive, social
Karcher (2005b)	54	4	0.31	Journal article	RCT	6	Moderate	Psychosocial	School, social
Sheehan et al. (1999)	62	3	0.96	Journal article	Quasi	18	Weak	Problem behavior	Cognitive, psychological
Smith (2011)	72	3	0.23	Journal article	RCT	2	Moderate	Health	Health
Tomlin (1994)	55	10	0.49	Dissertation	Quasi	2	Strong	Academic	School, psychological, cognitive, social

Abbreviations: # ES, number of effect sizes per manuscript; *N*, highest sample size included in analysis; Quasi, Quasi-experimental design; RCT, randomized controlled trial.

TABLE 3 Descriptive information for coded moderators

Moderator	Minimum	Maximum	Mean
Mentee characteristics			
Percentage male	31	100	55
Percentage White (non-Hispanic)	0	86	35
Percentage Black	9	100	50
Percentage Hispanic	0	39	15
Age	9	12.5	11
Sample size	26	416	114
Mentor characteristics			
Percentage male	39	100	61
Percentage White	0	94	48
Percentage Black	0	100	43
Percentage Hispanic	0	22	7
Mentor incentive	40% yes (4, 5), 60% no (2, 3, 6)		
Program characteristics			
Program length in months	2	18	8
Pre-match training hours	2	8	5
Meeting frequency	1	9	4
Program location	60% urban (3, 4, 6), 40% rural (2, 5)		
Primary focus	50% general (1, 2, 3), 17% academic (6), 17% prob. beh. (4), 17% health (5)		
Primary site	67% school, during/after (1, 2, 5, 6), 33% other (3, 4)		
Type of mentoring	33% one-to-one (1, 6), 67% combination one-to-one and group (2, 3, 4, 5)		
Sole or primary intervention	67% sole intervention (1, 2, 3, 6), 33% primary intervention (4, 5)		
Level of adult oversight/ supervision 80% moderate-high (2, 3, 5, 6), 20% low (1)	80% moderate-high (2, 3, 5, 6), 20% low (1)		
Family component	40% yes (2, 3), 60% no (4, 5, 6)		
Methodological characteristics			
Year of publication	1994 (6), 1999 (4), 2002 (3), 2005 (2), 2008 (1), 2011 (5)		
Publication status	67% published (2, 3, 4, 5), 17% program report (1), 17% dissertation (6)		
Study design	67% RCT (1, 2, 3, 5), 33% Quasi experimental (4, 6)		
Study quality	17% strong (6), 67% moderate (1, 2, 3, 5), 17% weak (4)		
Source	55% mentee, 33% teacher, 12% school records		

Note: Roman numerals in parentheses refer to the studies (see corresponding Roman numerals in Table 1) coded in each category.

size of studies ($n = 6$), moderators were only included in analyses if the moderator could be coded in at least 3 out of 6 of the studies (50%). Unfortunately, none of the coded match characteristics (i.e., percentages of cross-race and cross-sex matches; matching based on sex, race, or interests; whether mentors and youth came from the same communities; and average age difference between mentors and mentees) were able to be included in the analyses due to infrequent reporting of these variables across studies.

Mentee characteristics

Of the coded mentee characteristics, only sample size, gender, race, and average age were reported in at least 50% of the studies and were included in the analyses. The mentee race/ethnicities reported within the included studies were limited to White (non-Hispanic), Black, and Hispanic. There were no significant differences in study effect sizes based on these mentee characteristics.

Mentor characteristics

Of the coded mentor characteristics, gender, race, and mentor incentive could be coded in at least 50% of studies and therefore were included in the analyses. The mentor races reported within the included studies were limited to White (non-Hispanic), Black, and Hispanic. There were no differences in effect size based on mentor characteristics.

Program characteristics

All of the coded program characteristics were reported in at least 50% of studies and therefore were included in the analyses. Results showed that there were statistically significant differences in the impact of youth mentoring based on the primary site of the program ($F[1, 41] = 10.03$, $p < .01$), with programs operating in the community or outside of the school day (i.e., weekend or summer) yielding larger effects than programs that were school-based (i.e., in schools, during or after the school day; $B = -0.64$, $t = -3.17$, $p < .01$). Results also showed that there were differences in the impact of youth mentoring based on the geographical location of the program ($F[1, 22] = 4.26$, $p < .05$), with programs in urban locations yielding larger effects than programs in rural locations ($B = .43$, $t = 2.06$, $p < .05$). Additionally, results demonstrated a difference in the impact of cross-age peer mentoring based on the level of adult oversight/supervision of program interventions ($F[1, 38] = 4.91$, $p < .05$). Programs with moderate to high levels of adult oversight or supervision yielded larger effects than programs with low levels of adult oversight or supervision ($B = .4$, $t = 2.22$, $p < .05$).

TABLE 4 Moderators of the effectiveness of mentoring programs

Moderator variable	*k*	*#ES*	B_0/*g*	t_0	B_1	t_1	$F(df_1, df_2)$
Broad outcome domains							$F(4, 38) = 3.74$
School (RC)	4	18	0.42	2.50*			
Psychological outcomes	2	9	0.48	2.78**	0.06	0.94	
Health	2	4	0.59	3.18**	0.17	1.63	
Cognitive functioning	3	5	0.33	1.54	−0.09	−0.52	
Social	4	7	0.46	2.59**	0.03	0.47	
Narrow outcome sub-categories							
School							$F(1, 16) = 0.02$
Academic outcomes (RC)	4	11	0.40	2.21*			
School engagement	4	7	0.39	2.14*	−0.01	−0.14	
Psychological symptoms							$F(1, 7) = 3.14$
Externalizing (RC)	2	8	0.29	1.14			
Overall mental health	1	1	1.19	2.69**	0.90	1.77	
Health							$F(1, 2) = 0.03$
Substance use (RC)	1	1	0.20	1.89+			
Physical health	1	3	0.23	1.63	0.03	0.19	
Social functioning							$F(1, 5) = 1.48$
Social skills (RC)	1	2	0.04	0.51			
Overall social support	4	5	0.18	2.07*	0.14	1.22	
Mentee characteristics							
Percentage male	4	21	0.42	2.27*	0.002	0.21	$F(1, 19) = 0.04$
Race/Ethnicity							$F(2, 14) = 3.43$
Percent White non-Hisp (RC)	3	17	0.80	3.62***			
Percent Black	3	17	0.49	3.54***	−0.31	−1.26	
Percent Hispanic	3	17	1.04	2.50*	0.37	1.09	
Age	6	43	1.02	0.74	−0.05	−0.42	$F(1, 41) = 0.18$
Sample size	6	43	0.62	3.99***	−0.002	−1.83+	$F(1, 41) = 3.35$+
Mentor characteristics							
Percentage male	3	18	0.54	1.11	−0.04	−0.05	$F(1, 16) = 0.003$
Race/Ethnicity							$F(2, 15) = 2.77$
Percent White non-Hisp (RC)	3	18	0.24	1.16			
Percent Black	3	18	0.48	3.67***	0.24	1.00	
Percent Hispanic	3	18	2.71	1.96*	2.28	1.64	
Mentor incentive							$F(1, 22) = 0.04$
No (RC)	2	6	0.52	2.61**			
Yes	3	18	0.59	2.36*	0.07	0.21	
Program characteristics							
Program length in months	6	43	0.16	0.69	0.04	1.53	$F(1, 41) = 2.33$

TABLE 4 (Continued)

Moderator variable	*k*	*#ES*	B_0/*g*	t_0	B_1	t_1	$F(df_1, df_2)$
Pre-match training hours	3	17	0.55	3.53***	−0.04	−1.18	$F(1, 15) = 1.39$
Meeting frequency	5	43	0.73	3.16**	−0.07	−1.51	$F(1, 41) = 2.29$
Primary focus							$F(3, 39) = 2.16$
General (RC)	3	27	0.35	1.58			
Academic	1	10	0.48	1.30	0.13	0.30	
Problem behavior	1	3	0.95	2.43*	0.60	1.33	
Health	1	3	0.23	0.60	−0.12	−0.27	
Program location							$F(1, 22) = 4.26$*
Rural (RC)	2	7	0.27	1.69+			
Urban	3	17	0.70	5.33***	0.43	2.06*	
Primary site							$F(1, 41) = 10.03$**
Other (RC)	2	7	0.89	5.05***			
During/after school	4	36	0.25	2.47*	−0.64	−3.17**	
Type of mentoring							
Individual and group (RC)	4	14	0.56	3.15**			$F(1, 41) = 1.16$
Individual	2	29	0.25	1.08	−0.31	−1.08	
Sole or primary intervention							$F(1, 41) = 1.11$
Primary (RC)	2	4	0.82	2.16*			
Sole	4	39	0.39	2.54*	−0.43	−1.05	
Level of adult oversight							$F(1, 38) = 4.91$*
Low (RC)	1	19	0.03	0.23			
Moderate-High	4	21	0.43	4.28***	0.40	2.22*	
Family component							$F(1, 22) = 0.0004$
No (RC)	3	16	0.55	2.77**			
Yes	2	8	0.54	2.12*	−0.01	−0.02	
Methodological characteristics							
Year of publication	6	43	−0.10	−0.28	0.03	1.58	$F(1, 41) = 2.51$
Publication status							$F(2, 40) = 2.03$
Published journal article (RC)	4	14	0.56	3.11**			
Dissertation	1	10	0.48	1.42	−0.08	−0.22	
Report	1	19	0.03	0.10	−0.53	−1.42	
Study design							$F(1, 41) = 2.02$
RCT (RC)	4	30	0.31	1.96+			
Quasi-experimental	2	13	0.70	3.15**	0.39	1.42	
Study quality							$F(2, 40) = 3.32$
Strong (RC)	1	10	0.48	1.64			
Moderate	4	30	0.31	1.20*	−0.17	−0.52	
Weak	1	3	0.95	2.98**	0.47	1.09	

(Continues)

TABLE 4 (Continued)

Moderator variable	*k*	*#ES*	B_0/g	t_0	B_1	t_1	$F(df_1, df_2)$
Source							$F(2, 39) = 4.60^{+}$
Mentee (RC)	6	23	0.41	2.87**			
Teacher	3	14	0.42	2.88**	0.02	0.32	
School records	2	5	0.77	3.72***	0.36	2.13*	

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

Abbreviations: k, number of studies; #ES, number of effect sizes; RC, reference category; 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 reference category; $F(df_1, df_2)$ omnibus test.

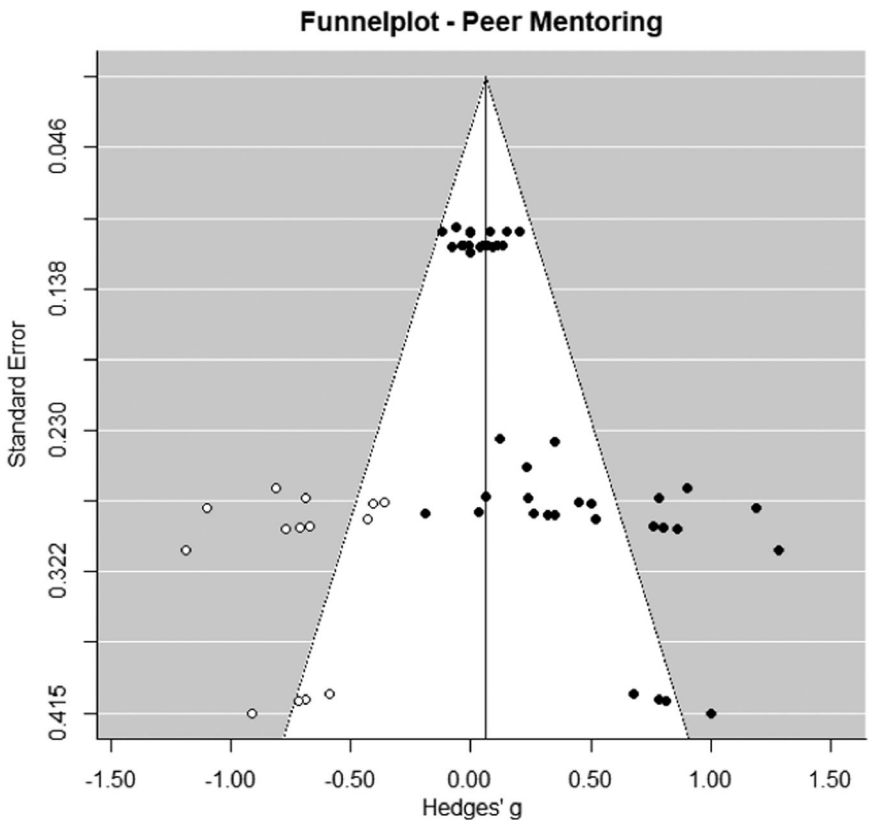

FIGURE 2 Funnel plot analysis

There were no differences in the impact of youth mentoring programs based on expected program length, meeting frequency, number of pre-match training hours, or program primary focus. Likewise, no moderation was observed for the type of mentoring, whether there was a family component, or whether mentoring was the sole or primary intervention in the program.

Methodological characteristics

All of the identified methodological characteristics could be coded in at least 50% of studies and therefore were included in the analyses. Results showed that there was significant variability in effect sizes across source of outcome information ($F[1, 40] = 4.50$, $p < .05$), with school records ($B = 0.36$, $t = 2.12$, $p < .05$) yielding larger effect sizes than effect sizes calculated from other assessment methods (i.e., youth self-report, parent report, and teacher report). There were no significant differences in effect sizes based on year of study publication, publication status, study design, or ratings of overall study quality.

PUBLICATION BIAS AND SENSITIVITY ANALYSES

Funnel plot analysis showed some indication of publication bias. Fourteen effect sizes were missing at the left side of the of the funnel plot (see Figure 2). Accounting for publication bias by means of a trim and fill analysis yielded a smaller nonsignificant mean effect size of Hedges'

TABLE 5 Leave-one-out sensitivity analyses

Overall effect	# studies	# ES	Effect size (SE)	CI 95%	*p*
Cross-age youth mentoring programs	6	43	0.45 (0.14)	(0.16–0.74)	.003**
- Excluding (Herrera et al., 2008)	5	24	0.54 (0.13)	(0.26–0.82)	<.001**
- Excluding (Karcher, 2005b)	5	39	0.48 (0.17)	(0.13–0.83)	.009**
- Excluding (Sheehan et al., 1999)	5	40	0.34 (0.12)	(0.09–0.59)	.009**
- Excluding (Smith, 2011)	5	40	0.49 (0.17)	(0.15–0.83)	.006**
- Excluding (Karcher et al., 2002)	5	39	0.39 (0.15)	(0.08–0.69)	.015*
- Excluding (Tomlin, 1994)	5	33	0.45 (0.18)	(0.08–0.81)	.018*

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

Abbreviations: CI, confidence interval # ES, number of effect sizes; mean g, mean effect sizes; SE, standard error; # studies, number of studies.

$g = 0.19$ ($p = .36$). However, as noted above, there is still no valid and reliable way to examine publication bias in multilevel meta-analysis. Moreover, the funnel plot method assumes homogeneity of the overall effect size, an assumption which was violated in this study. Therefore, this finding should be interpreted with caution.

Findings from the leave-one-out sensitivity analyses (see Table 5) indicated that the overall effect remained significant after each rerun; therefore, none of the studies had an individual, disproportionate, impact on the overall findings. Moreover, the interval of effect sizes obtained through the sensitivity analyses ($0.34 < g < 0.54$) contains the overall effect size based on the total set of studies (0.45) and overlaps with the 95% confidence interval of the total effect size (95% CI: 0.17–0.73).

DISCUSSION

This study represents the first comprehensive meta-analysis of cross-age peer mentoring programs for youth. Analyses revealed a medium-sized overall effect of cross-age peer mentoring programs ($g = 0.45$), with no differences in mentoring impact across different types of youth outcomes. Several program and methodological characteristics did moderate effect sizes, with larger effects for programs that were operated in the community or outside of the school day (i.e., weekend or summer), conducted in urban settings, and included moderate to high levels of adult oversight and supervision. Although these findings should be interpreted with caution given the limited number of programs included in this study, the results highlight the potential benefits of cross-age peer mentoring for youth and indicate a need for further research in this area.

This meta-analysis builds upon past meta-analyses that have demonstrated the impact of intergenerational youth mentoring programs in which an adult mentor is paired with the youth mentee, with many of those studies yielding statistically significant, but small effect sizes (i.e., ranging from $g = 0.18$–0.21; DuBois et al., 2002, 2011; Raposa et al., 2019). The effect size for cross-age peer mentoring in our study is more than double that observed in these past meta-analyses. This difference is notable, given the far greater emphasis on intergenerational mentoring programs compared to cross-age peer mentoring programs in practice and in the research literature. Cross-age peer mentoring can offer feasible and efficient opportunities to have older peers mentor youth (e.g., by pairing up youth from different grades within the same school) with the potential for mutual benefit. It is also important to note the limited sample size ($n = 6$) of studies in our analyses, which likely influenced the lack of differentiation in effect sizes across different types of youth outcomes. This small sample size is due, in part, to rigorous inclusion criteria that adhered strictly to a cross-age peer mentoring framework, such that only studies that evaluated mentoring programs in which an older youth (at least two years older) acted in a non-professional helping capacity with a specific younger person to promote positive youth outcomes were included (Karcher & Berger, 2017). Nevertheless, future studies should replicate and expand upon these results when analyses may be better powered to detect heterogeneity across effect sizes and moderators that explain this heterogeneity.

Moderator analyses also suggested a number of factors that may influence the impact of cross-age peer mentoring for youth. Consistent with some previous findings (e.g., Durlak, 1979; Herrera et al., 2008), programs with moderate to high levels of adult oversight and supervision had larger effects compared to programs with low levels of adult oversight and supervision. Elements of strong adult oversight and supervision included: mandatory training for mentors; supervision to support intervention delivery; videotaping select mentor-mentee interactions to monitor intervention quality and provide additional support as

needed; and program staff and parent participation in program activities (Karcher, 2005b; Karcher et al., 2002; Smith, 2011; Tomlin, 1994). In contrast, programs with low adult oversight and supervision provided less than two hours of training, with a significant portion of youth mentors reporting receiving no training at all (Herrera et al., 2008).

Adult oversight and supervision may improve treatment fidelity, a construct that has repeatedly been linked to better treatment outcomes (Durlak & DuPre, 2008; Schoenwald et al., 2000). Adult supervision provides opportunities for verifying that mentoring interventions are carried out as intended, while also ensuring appropriate mentor-mentee interactions and preventing obstacles to program success (e.g., mentors not understanding their role or carrying out their role ineffectively). This oversight may be particularly important for youth mentors, given their developmental stage, maturity level, and the potential for negative peer influence (Dishion & Tipsord, 2011; Karcher, 2008; O'Donnell et al., 1979). Additional scaffolding for younger mentors may facilitate clearer expectations of their roles and responsibilities, as well as ensuring adherence to program interventions.

Results also demonstrated that peer mentoring programs in urban settings yielded larger effects than those in rural settings. Further research is needed to replicate and explore this finding before firm conclusions can be drawn. However, it is possible that programs in urban settings may serve youth who experience increased environmental risk factors such as exposure to high rates of crime, violence, delinquency, substance use, and poverty (Black & Krishnakumar, 1998). Although specific individual and environmental risk factors were not reported in most studies in this meta-analysis, the three studies serving youth in urban settings in our sample included, (1) a mentoring program aimed at violence prevention among youth living in a violent neighborhood; (2) a program aimed to improve academic attainment of black adolescent males with high risk ratings; and (3) a program serving youth in school districts with the highest dropout rates in the city (Sheehan et al., 1999, Tomlin, 1994, and Karcher et al., 2002, respectively). Some meta-analytic evidence from intergenerational mentoring suggests that programs serving youth with greater levels of these kinds of environmental risk factors may tend toward larger effect sizes (DuBois et al., 2011). This study supports the relevance of the "pendulum" theory of change for youth mentoring programs, which states that more vulnerable youth have the most room for improvement (Tanner-Smith et al., 2018). Future research should continue to assess whether youth with greater environmental risk (across both urban and rural settings) may be especially likely to benefit from cross-age peer mentoring interventions.

Additional moderators of program effect size were also observed. Larger effects were observed for programs operating in the community or outside of the school day (i.e., weekend or summer) compared to those that were school-based (i.e., in schools, during or after the school day). It is possible that space and time constraints within school settings may restrict the types of activities in which matches can engage. School-based programs operating during or after the school day may be more likely to meet within a group format in a space like the school cafeteria, which could distract from intended one-on-one interventions (e.g., Herrera et al., 2008; Karcher, 2005b; Smith, 2011; Tomlin, 1994). In addition, school-based programs that meet after the school day ends (e.g., Herrera et al., 2008; Karcher, 2005b; Smith, 2011) may be negatively affected by youth fatigue that limits mentees' ability to fully engage in program content, as well as mentors' capacity to deliver interventions successfully. Future research should replicate and extend these findings, perhaps exploring the interacting factors that may influence the impact of program setting on effect size. This area of research is particularly important given that many youth access older peer mentorship within school settings, and school-based mentoring has become one of the most popular contexts in which youth receive mentoring services (Garringer et al., 2017).

Finally, inconsistent with prior meta-analyses of intergenerational mentoring and other psychosocial interventions (A. Cheung & Slavin, 2015; Raposa et al., 2019), this study found that school records (e.g., GPA, school absences, suspensions, and office referrals) yielded higher effect sizes compared to mentee and teacher self-report measures. These results may reflect the effectiveness of cross-age peer mentoring in influencing specific behavioral and academic outcomes (Karcher et al., 2002; Tomlin, 1994). These findings may also suggest that cross-age peer mentoring interventions are less effective means for changing mentee attitudes and teachers' perceptions of their students, or that these changes are not well captured by the included measures. It is possible that mentees' school-related behavior may improve without corresponding attitudinal changes. Future research should continue to explore these possibilities, and 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 mentoring program effectiveness.

A final set of analyses explored the impact of publication bias on these findings, as well as the influence of outliers within the study sample on the findings. Leave-one-out sensitivity analyses indicated that the overall impact of cross-age peer mentoring remained similar after each rerun, indicating that none of the studies had a disproportionate impact on the overall findings. However, funnel plot analysis showed some indication of publication bias, with 14 effect sizes missing at the left side of the of the funnel plot, and accounting for publication bias by means of a trim and fill analysis yielded a smaller and non-significant mean effect size. Given that the funnel plot method was not originally developed for multilevel meta-analyses, and that it assumes homogeneity of the overall effect size—an assumption which was violated in this s tudy—this finding should be interpreted with caution. Nevertheless, this result points to the need for additional

evaluation studies of cross-age peer mentoring programs for youth, as well as meta-analytic efforts to replicate the present findings.

There are several limitations to the current meta-analysis. First, meta-analyses are dependent on the type, quality, and availability of information included in the analyses. Within the studies that were included, for example, half did not report mentee or mentor racial background, only one study had a large sample size ($n > 100$), and certain moderators could not be included in the analysis due to lack of reporting (e.g., how youth were matched with mentors). As noted previously, only six studies met the rigorous inclusion criteria for this study, which were based in Karcher and Berger's (2017) definition of cross-age peer mentoring. By limiting the inclusion criteria to fit this definition, this study aimed to fill a specific gap in the literature, investigating the impact youth mentors have on younger peers and determining the effectiveness of cross-age peer interventions within the mentoring field. Nevertheless, these findings should be interpreted with caution given the small sample size of the present meta-analysis. As additional cross-age peer mentoring programs are implemented and evaluated, a follow-up meta-analysis should further evaluate overall effectiveness of cross-age peer mentoring programs, as well as moderator variables that impact mentee outcomes.

While the present analyses included many methodological factors as moderators of study outcomes, issues such as reporter bias or unreliable or poorly-validated measurement tools in the original studies may also have impacted observed effect sizes for these studies. In addition, certain moderators identified as potentially relevant based on previous research or theory could not be tested because of the lack of consistent reporting of these variables. As this literature grows, future meta-analyses should analyze additional relevant moderators from the mentoring literature and investigate complex interactions between moderating variables (Raposa et al., 2019). This is particularly important given the substantial heterogeneity observed across the studies included in the current analyses (and indeed observed in most multi-level meta-analyses of youth mentoring programs; e.g., Raposa et al., 2019).

Finally, only evaluations written in English were coded and included in the analysis given limitations of the language proficiency of this study's coders and authors. Although most studies screened were written in English, some evaluations were excluded based purely on language of publication. Moreover, the studies that met criteria for inclusion were all conducted in the United States. Therefore, the present findings may not generalize to cross-age peer mentoring programs outside of the United States. Future research should explore the various factors that may influence cross-age peer mentoring programs in other countries and contexts, including cultural and demographic variables.

Despite these limitations, this study provides the first meta-analytic assessment of the overall impact of one-on-one, cross-age peer mentoring relationships, as well as moderators of these programs' effectiveness. Taken together, the current findings provide support for the efficacy of programs that foster one-on-one, caring relationships between youth and older peers who are closely supervised by adults, particularly as an intervention with the potential to reach large groups of youth and prevent more intensive treatments. Future research should continue to include rigorous, experimental evaluations of cross-age peer mentoring programs, including measurement of youth, mentor, match, and program characteristics that may influence program impact. At the same time, these findings highlight opportunities for improving the quality and rigor of mentoring practice, particularly moving toward interventions with strong attention to oversight of mentors. Mentoring programs should endeavor to build on the current findings by creating supportive, scaffolding environments for young mentors, thereby facilitating increased mentor competency and fostering mentee success. Increased implementation and evaluation of cross-age peer mentoring programs in this way is a promising path to scale supplemental and preventative services to youth.

REFERENCES

Adler, A. (1964). *Problems of neurosis*. Harper Torchbooks.

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*(3), 154–174. https://doi.org/10.20982/tqmp.12.3.p154

Black, M. M., & Krishnakumar, A. (1998). Children in low-income, urban settings: Interventions to promote mental health and well-being. *American Psychologist*, *53*(6), 635–646. https://doi.org/10.1037/0003-066X.53.6.635

Bocknek, E. L., Sanderson, J., & Britner, P. A., IV (2009). Ambiguous loss and posttraumatic stress in school-age children of prisoners. *Journal of Child and Family Studies*, *18*(3), 323–333. https://doi.org/10.1007/s10826-008-9233-y

Cheung, A., & Slavin, R. E. (2015). How methodological features affect effect sizes in education, *Best evidence encyclopedia* (pp. 1–35). Johns Hopkins University. http://www.bestevidence.org/word/methodological_Sept_21_2015.pdf

Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, *19*(2), 211–229. https://doi.org/10.1037/a0032968

Christensen, K. M., Hagler, M. A., Stams, G., Raposa, E. B., Burton, S., & Rhodes, J. E. (2020). Non-specific versus targeted approaches to youth mentoring: A follow-up meta- analysis. *Journal of Youth and Adolescence*, *49*(5), 959–972. https://doi.org/10.1007/s10964-020-01233-x

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

Dishion, T. J., & Tipsord, J. M. (2011). Peer contagion in child and adolescent social and emotional development. *Annual Review of Psychology*, *62*, 189–214. https://doi.org/10.1146/annurev.psych.093008.100412

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. (1979). Comparative effectiveness of paraprofessional and professional helpers. *Psychological Bulletin*, *86*(1), 80–92. https://doi.org/10.1037/0033-2909.86.1.80

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*(3–4), 327–350. https://doi.org/10.1007/s10464-008-9165-0

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*(2), 455–463. https://doi.org/10.1111/j.0006-341x.2000.00455.x

Ellis, S., Rogoff, B., & Cromer, C. C. (1981). Age segregation in children's social interactions. *Developmental Psychology*, *17*(4), 399–407. https://doi.org/10.1037/0012-1649.17.4.399

Feldman, R. A. (1992). The St. Louis experiment: Effective treatment of antisocial youths in prosocial peer groups. In J. McCord, & R. E. Tremblay (Eds.), *Preventing antisocial behavior: Interventions from birth through adolescence* (pp. 232–252). Guilford.

Garringer, M., McQuillin, S., & McDaniel, H. (2017). *Examining youth mentoring services across America: Findings from the 2016 National Mentoring Program survey*. MENTOR: The National Mentoring Partnership. http://www.mentoring.org/new-site/wpcontent/uploads/2017/07/Mentor-Survey-Report_FINAL_small.pdf

Harris, J. R. (1995). Where is the child's environment? A group socialization theory of development. *Psychological Review*, *102*(3), 458–489. https://doi.org/10.1037/0033-295X.102.3.458

Harris, J. R. (1998). *The nurture assumption: Why children turn out the way they do?* Free Press.

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

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*. Public/Private Ventures.

Hirschi, T. (1969). *Causes of delinquency*. University of California Press.

Jessor, R., & Jessor, S. L. (1977). *Problem behavior and psychosocial development: A longitudinal study of youth*. Academic Press.

Karcher, M. J. (2005a). Cross-age peer mentoring. *Handbook of Youth Mentoring*, 233–257.

Karcher, M. J. (2005b). The effects of school-based developmental mentoring and mentors' attendance on mentees' self-esteem, behavior, and connectedness. *Psychology in the Schools*, *42*, 65–77. https://doi.org/10.1002/pits.20025

Karcher, M. J. (2008). The cross-age mentoring program (CAMP): A developmental intervention for promoting students' connectedness across grade levels. *Professional School Counseling*, *12*(2), 137–143. https://doi.org/10.1177/2156759X0801200208

Karcher, M. J., & Berger, J. R. M. (2017). *One-to-one cross-age peer mentoring*. National Mentoring Resource Center (NMRC), Office of Juvenile Justice and Delinquency Prevention (OJJDP). https://nationalmentoringresourcecenter.org/index.php/what-works-in-mentoring/model-and-population-reviews.html?layout=edit&id=301

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

Karcher, M. J., Hansen, K., Herrera, C., & Crisp, G. (2011). *The right focus: How mentor, mentee and program characteristics contribute to program and relationship quality*. Poster presented at the 2011 Biennial Meeting of the Society for Research in Child Development. Montreal, Quebec, Canada.

Karcher, M. J., & Lindwall, J. (2003). Social interest, connectedness, and challenging experiences. What makes high school mentors persist? *Journal of Individual Psychology*, *59*, 293–315. http://adolescentconnectedness.com/media/KarcherLindwall_teenmentors.pdf

Karcher, M. J., Nakkula, M. J., & Harris, J. (2005). Developmental mentoring match characteristics: The effects of mentors' efficacy and mentees' emotional support seeking on the perceived quality of mentoring relationships. *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), 2693–2627. https://doi.org/10.1002/sim.1482

Lau, J., Ioannidis, J. P., Terrin, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. *British Medical Journal*, *333*, 597–600. https://doi.org/10.1136/bmj.333.7568.597

Lipsey, M. (2006). The effects of community-based group treatment for delinquency: a meta- analytic search for cross-study generalizations. In K. A. Dodge, T. J. Dishion, & J. E. Lansford (Eds.), *Deviant peer influences in programs for youth: Problems and solutions* (pp. 162–184). Guilford.

McCord, J. (2003). Cures that harm: Unanticipated outcomes of crime prevention programs. *The Annals of the American Academy of Political and Social Science*, *587*(1), 16–30. https://doi.org/10.1177/0002716202250781

Miller-Johnson, S., & Costanzo, P. (2004). If you can't beat 'em… induce them to join you: Peer-based interventions during adolescence. In J. B. Kupersmidt, & K. A. Dodge (Eds.), *Children's peer relations: From development to intervention* (pp. 209–222). American Psychological Association. https://doi.org/10.1037/10653-011

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

O'Donnell, C. R., Lydgate, T., & Fo, W. S. (1979). The Buddy System: Review and follow-up. *Child Behavior Therapy*, *1*(2), 161–169. https://doi.org/10.1300/J473v01n02_03

O'Donnell, C. R., & Williams, I. L. (2013). The buddy system: A 35-year follow-up of criminal offenses. *Clinical Psychological Science*, *1*(1), 54–66. https://doi.org/10.1177/2167702612456907

Poon, C. Y. S. (2019). "*It's more complicated than that!": The influence of multiple presenting problems on youth outcomes in formal mentoring programs* [Masters' thesis, University of Massachusetts Boston]. ProQuest Dissertations & Thesis Global.

Raposa, E. B., Rhodes, J., Stams, G. J., Card, N., Burton, S., Schwartz, S., YovieneSykes, L. A., Kanchewa, S., Kupersmidt, J., & Hussain, S. (2019). The effects of youth mentoring programs: A meta-analysis of outcome studies. *Journal of Youth and Adolescence*, *48*, 423–443. https://doi.org/10.1007/s10964-019-00982-8

Rhodes, J. E. (2020). *Older and wiser: New ideas for youth mentoring in the 21st century*. Harvard University Press.

Santrock, J. (2019). *Adolescence*. McGraw-Hill Education.

Schoenwald, S. K., Henggeler, S. W., Brondino, M. J., & Rowland, M. D. (2000). Multisystemic therapy: Monitoring treatment fidelity. *Family Process*, *39*(1), 83–103. https://doi.org/10.1111/j.1545-5300.2000.39109.x

Selman, R. L. (1980). *The growth of interpersonal understanding*. Academic Press.

Sheehan, K., DiCara, J. A., LeBailly, S., & Christoffel, K. K. (1999). Adapting the gang model: Peer mentoring for violence prevention. *Pediatrics*, *104*(1), 50–54.

Smith, L. H. (2011). Piloting the use of teen mentors to promote a healthy diet and physical activity among children in Appalachia. *Journal for Specialists in Pediatric Nursing*, *16*(1), 16–26.

Stevens, E. A., & Prinstein, M. J. (2005). Peer contagion of depressogenic attributional styles among adolescents: A longitudinal study. *Journal of Abnormal Child Psychology*, *33*(1), 25–37. https://doi.org/10.1007/s10802-005-0931-2

Stukas, A. A., & Tanti, C. (2005). Recruiting and sustaining volunteer mentors. *Handbook of Youth Mentoring*, *33*, 235–250.

Tabachnik, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). 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. https://doi.org/10.1007/s11121-018-0942-1

Tomlin, V. E. (1994). *A mentor program for improving the academic attainment of black adolescent males* (unpublished doctoral dissertation). University of Denver, Colorado.

Treadwell, J. R., Tregear, S. J., Reston, J. T., & Turkelson, C. M. (2006). A system for rating the stability and strength of medical evidence. *BMC Medical Research Methodology*, *6*, 52–72. https://doi.org/10.1186/1471-2288-6-52

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*(4), 1274–1294. https://doi.org/10.3758/s13428-014-0527-2

Vázquez, A. L., & Villodas, M. T. (2019). Racial/ethnic differences in caregivers' perceptions of the need for and utilization of adolescent psychological counseling and support services. *Cultural Diversity and Ethnic Minority Psychology*, *25*(3), 323–330. https://doi.org/10.1037/cdp0000255

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta- analysis. *Research Synthesis Methods*, *1*(2), 112–125. https://doi.org/10.1002/jrsm.11

What Works Clearinghouse Review of the Report. (2013). *High school students as mentors: Findings from the Big Brothers Big Sisters school-based mentoring impact study*. Institute of Education Sciences, United States Department of Education.

Wheeler, M. E., Keller, T. E., & Dubois, D. L. (2010). Review of three recent randomized trials of school-based mentoring. *Social Policy Report*, *24*(3), 1–21. https://core.ac.uk/download/pdf/37774968.pdf

How to cite this article: Burton, S., Raposa, E. B., Poon, C. Y. S., Stams, G. J. J. M., & Rhodes, J. (2021). Cross-age peer mentoring for youth: A meta-analysis. *American Journal of Community Psychology*, 1–17. https://doi.org/10.1002/ajcp.12579