



# Effects of the Big Brothers Big Sisters of America Community-Based Mentoring Program on social-emotional, behavioral, and academic outcomes of participating youth: A randomized controlled trial

Carla Herrera<sup>a,\*</sup>, David L. DuBois<sup>b</sup>, Janet Heubach<sup>c</sup>, Jean B. Grossman<sup>d,e</sup>

<sup>a</sup> Herrera Consulting Group, LLC, 6123 Utah Ave. NW, Washington, DC 20015, USA

<sup>b</sup> Institute for Health Research and Policy (M/C 275), University of Illinois Chicago, 1747 W. Roosevelt Rd., Chicago, IL 60608, USA

<sup>c</sup> MENTOR Washington, 15500 SE 30th Pl #200, Bellevue, WA 98007, USA

<sup>d</sup> MDRC, 200 Vesey Street 23rd Floor, New York, NY 10281–2103, USA

<sup>e</sup> Princeton School of Public and International Affairs, Princeton University, Robertson Hall, 20 Prospect Ave, Princeton, NJ 08540, USA

## ARTICLE INFO

### Keywords:

Mentoring  
Evaluation  
Mental health  
Positive youth development

## ABSTRACT

This research is a randomized controlled trial of effects of the Big Brothers Big Sisters of America (BBBSA) Community-Based Mentoring Program on the social-emotional, behavioral, and academic outcomes of participating youth over a 13-month period. Seven hundred sixty-four youth between 9 and 14 years old were enrolled in the study through two BBBSA agencies that served predominantly urban areas on the West Coast. Each participating youth was randomly assigned either to be immediately eligible for being matched with a mentor (treatment group;  $n = 379$ ) or to remain on the program waitlist for 13 months (control group;  $n = 385$ ). Analyses for the present study are based on 654 youth for whom outcome data were able to be obtained at a 13-month follow-up (87.0% of the treatment group and 84.2% of the control group). Outcome measures were completed by youth (depressive symptoms, prosocial behavior, social acceptance, parent-child relationship quality, misconduct, self-perception of academic ability, skipping school, and academic performance) and primary caregivers (Strengths and Difficulties Questionnaire [SDQ]) at baseline and at the 13-month follow-up. Composite indices reflecting the average of youth- and/or parent-reported outcome measures were also examined. Findings indicated effects of mentoring (i.e., being offered a mentor) that reached statistical significance, all favoring the treatment group, for youth-reported depressive symptoms, parent-reported emotional symptoms, peer problems, conduct problems and the total difficulties score on the SDQ, and the parent-report and combined youth- and parent-report composite indices at the 13-month assessment (Cohen's  $d$  ranging from 0.138 to 0.253).

## 1. Introduction

Mentoring programs are one of the most widely used forms of support for youth from socio-economically disadvantaged backgrounds. Meta-analyses of extant evaluations of mentoring programs have supported their potential to promote improved outcomes for participating youth in areas such as academics, social competence, and problem behavior avoidance (for reviews, see DuBois et al., 2002; DuBois et al., 2011; Tolan et al., 2014; Raposa et al., 2019). Tolan et al. (2014), for example, found standardized mean difference estimated effect sizes of 0.11 for outcomes relating to academic achievement, and 0.21, 0.29, and 0.16 for those relating to delinquency, aggression, and drug use, respectively.

Yet, the mentoring programs in the field today (and on which these

meta-analyses are based) vary considerably along a number of salient dimensions (Garringer et al., 2017), all of which may have implications for program impacts. These include the type and scope of youth outcomes that are of primary interest, with some programs focused on improvements in a particular domain, such as academic achievement or delinquency prevention, and others seeking to foster positive development more broadly across multiple domains. Other noteworthy variations include program format (e.g., one-to-one versus group), mentors' age (e.g., older peer versus adult), how the mentor role is structured (e.g., volunteer versus paid paraprofessional), the nature of the activities mentors and youth are expected to engage in together, whether activities are limited to a particular setting (e.g., school), and whether additional components other than mentoring are included in the program model (Garringer et al., 2017). This programmatic diversity

\* Corresponding author at: Herrera Consulting Group, LLC, 6123 Utah Ave. NW, Washington, DC 20015, USA.

<https://doi.org/10.1016/j.childyouth.2022.106742>

Received 6 February 2022; Received in revised form 7 September 2022; Accepted 23 November 2022

Available online 28 November 2022

0190-7409/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

underscores the importance of generating rigorous evidence that addresses the effectiveness of specific programs and approaches. Program-specific findings are important not only for developing the field's knowledge base, but also for supporting the decision-making of practitioners, funders and policymakers.

The focus of the current research is the community-based mentoring (CBM) program of Big Brothers Big Sisters of America (BBBSA). In 2020, a total of 109,254 youth were served by BBBSA through over 230 agencies nationwide (Porzig, 2021). In BBBSA's CBM program model, participating youth are matched one-on-one with adult volunteers (DuBois & Friend, 2017). Each "match," comprised of a youth (Little Brother or Sister—"Little") and adult (Big Brother or Sister—"Big"), spends time together on a regular basis (e.g., a few times a month for a few hours each time) in recreational and social activities in locations of their choosing (DuBois & Friend, 2017). Program staff assume a number of important roles, including recruiting, screening and training mentors, matching youth and mentors based on considerations such as shared interests, providing ongoing support for relationships through regular check-ins with mentors and families, and supporting a closure process when matches end (DuBois & Friend, 2017). The aim of the BBBSA CBM program is to promote the overall positive development of participating youth in areas such as academic achievement, self-esteem, and avoidance of substance use and other risky behaviors.

The mentor-youth relationship, and the interactions that comprise it, are "front and center" in BBBSA CBM. This programmatic emphasis on the mentoring relationship is consistent with the theoretical model of youth mentoring developed by Rhodes (2005). This model emphasizes how mentoring interactions can provide youth with a positive, supportive role model and opportunities to develop new skills and ways of thinking that, in turn, support identity development as well as social-emotional and cognitive growth. These processes are assumed to be dependent, in large part, on the development of a strong, trusting relationship between the mentor and youth (Rhodes, 2005). Although the mentoring relationship itself is important in other types of mentoring programs, other aspects of mentor-youth interactions—for example, academic activities in programs specifically targeting academic outcomes (Larose & Tarabulsy, 2005) and peer interactions in group programs (Kuperminc & Thomason, 2014)—may also be of central importance and foster distinct outcomes.

In 1995, Public/Private Ventures (P/PV) published a landmark multi-site randomized controlled trial (RCT) evaluation of the BBBSA CBM program that involved 959 youth (Tierney et al., 1995). The study found evidence of significant favorable effects of program participation at an 18-month follow-up on a number of outcomes as reported by youth, including school grades, parental trust, fighting (hitting someone), skipping school, and initiation of drug use. However, there were a number of outcomes without evidence of significant impact and, on average, effect sizes were small in magnitude (standardized mean difference = 0.06; calculated by Herrera et al., 2007). A small number of studies have subsequently evaluated the effectiveness of the BBBSA CBM program, but they have had a variety of significant methodological limitations, such as quasi-experimental designs (DuBois et al., 2002) or high rates of attrition (ICF International, 2011). Likewise, although relatively small-scale randomized trials have been carried out on the CBM program as implemented by BBBS organizations in Canada (n = 71; DeWitt et al., 2007) and Ireland (n = 164; Dolan, Brady, O'Regan, Russell et al., 2010), both cultural and programmatic differences make it unclear as to what extent the findings of these studies can be generalized to the BBBSA CBM program as delivered in the U.S.

Also noteworthy are several important questions about the program which the P/PV study and existing evaluations do not address. First, the P/PV study assessed outcomes using only youth self-reported data. The findings thus leave open the question of whether the profile of the program's effects would be similar using additional sources of data, such as parent reports. Illustratively, a recent RCT of the *Friends of the Children* mentoring program found that program effects were evident on

several parent-reported measures, but none of the measures gathered directly from youth (Eddy et al., 2017). Second, although the P/PV study examined potential effects of the program across a relatively broad range of outcomes, recent research suggests a number of additional areas in which it would be useful to have a better understanding of the program's potential benefits. Some research, for example, points toward potential effectiveness of mentoring in reducing levels of emotional difficulties among youth (for a recent meta-analysis, see Claro & Perlmutter, 2021). Similarly, although not well studied, mentoring may also promote character development and prosocial behavior in youth through a variety of processes, such as mentor guidance and role modeling, mutual engagement in constructive activities such as community service, and decreases in rejection sensitivity (Kanchewa et al., 2018).

Third, with few exceptions, evaluations of mentoring programs to date have only assessed distinct youth outcomes separately from one another. This approach does not address the question of the level of benefit mentoring may have for youth when considering impact across a number of outcomes. It has been noted in literature on other forms of intervention (e.g., drug treatment) that effect sizes for a composite measure of all outcomes may be greater than not only the average effect size for the constituent outcome measures but also, in some circumstances, the maximum of the outcome-specific effect sizes (Liu-Seifert et al., 2017). These possibilities are more likely in the case of outcome measures that are not highly correlated (Liu-Seifert et al., 2017). Incorporating tests of effectiveness that are sensitive to such possibilities could be especially important for mentoring programs such as BBBSA CBM in which multi-faceted developmental gains are a goal, and outcomes of interest may tend to not be highly correlated (e.g., academic achievement and emotional well-being). Furthermore, practitioners and policymakers may find it informative to have an index of overall improvement that can be expected across differing areas if youth are provided access to a mentoring program. In line with these considerations, the authors of a recent randomized controlled trial of the *Rock Your Life!* mentoring program in Germany (Resnjanskij et al., 2021) constructed an index that combined measures of math grades, patience/social skills, and labor-market orientation (e.g., career knowledge) and found that youth assigned to the program scored significantly higher on this composite at follow-up relative to controls.

Fourth, the P/PV impact study provided limited information on implementation quality. In mentoring, key implementation practices include provision of initial orientation/training for mentors, ongoing monitoring and support contacts of staff with mentors and families, and the frequency and duration of mentor outings with youth. Implementation data in areas like these that are integral to a program's delivery model can provide a valuable context for interpreting program impacts, such as the extent to which weak or null effects may be attributable to inconsistencies in program delivery (Durlak & Dupre, 2008).

Finally, both the program and societal context in which the BBBSA CBM program operates have continued to evolve in ways that could affect the generalizability of past findings to the program's current effectiveness. For example, BBBSA has added children with incarcerated parents as one of its priority populations to serve (DuBois & Friend, 2017)—a change with potentially significant implications for the risk profile of youth served in the program. Practices, furthermore, can also vary significantly across agencies within the parameters of national BBBSA standards (Wheeler & DuBois, 2009). It is not uncommon, for example, for agencies to introduce additional training or support processes for mentors in response to local needs or funding opportunities. It will be important to understand how these changes may have affected program outcomes.

Several of these questions and issues have been addressed in recent rigorous evaluations of other mentoring programs (see, for example, Axford et al., 2021; Guryan et al., 2021; Heppel et al., 2021; King et al., 2021; Millenky et al., 2014), as well as of other BBBS program models/

partnerships (e.g., [Herrera et al., 2007](#); [Lindstrom Johnson et al., 2022](#)). However, to date, they have not been fully addressed in evaluations of BBBSA CBM—a mentoring program model with one of the largest reaches in the U.S.

The current study is a randomized controlled trial of the effectiveness of the BBBSA CBM program as implemented in two of its large West Coast agencies. Several features of the study are noteworthy and complement existing studies of the program. First, in line with aforementioned considerations, youth outcomes are assessed using both youth- and parent-reported measures.

Second, outcomes encompass multiple domains, including emotional well-being, prosocial behavior, social relationships, problem behavior and academics—many of which have not been included in previous BBBSA evaluations—and are analyzed both individually and in the aggregate to afford enhanced sensitivity to potential patterns of effects. Depressive symptoms, for example, have not been included as an outcome in prior evaluations of BBBSA CBM. Yet, the program may have strong potential to affect this important outcome. Increasing opportunities for youth to engage in activities they enjoy is a central focus of both BBBSA CBM and behavioral activation, an evidence-supported treatment for depression ([Cuijpers et al., 2007](#)). Emotional symptoms (e.g., anxiety, worrying) also have been excluded from past studies of BBBSA CBM. Yet, mentors could be instrumental in reinforcing adaptive thinking patterns and pointing out cognitive errors that can be linked with anxiety and other internalizing difficulties ([Kerr & King, 2014](#)). Also investigated are other outcomes that were included in the original P/PV study, albeit only through youth report, such as prosocial behaviors and conduct problems. Mentors may affect these outcomes by modeling positive behaviors, discouraging negative behaviors, and providing an interactive context in which to practice newly learned skills ([Rhodes, 2005](#)). Increasing prosocial behavior may, in turn, improve the quality of youth's relationships with both peers and other adults. Finally, academic outcomes (i.e., academic performance, skipping school, self-perception of academic ability) are also included in the current study as was the case in the original P/PV study. These outcomes could improve through a number of mechanisms including direct support from the mentor (e.g., discussing academics, homework help; see [Herrera et al., 2007](#); [Herrera et al., 2000](#)) and indirect mechanisms such as improvements in parental relationships ([Rhodes et al., 2000](#)).

Third, program implementation data were collected throughout the study from staff and encompass both activities led by staff and key information about the mentoring relationships (e.g., frequency of mentor-youth outings and relationship duration). Relative to national standards and typical affiliate requirements, each participating agency was asked to adopt somewhat more intensive practices over the course of the study in the areas of mentor training and ongoing support as well as increased expectations for frequency of contact and duration within their mentoring relationships. The study was thus conceptualized as potentially more likely to yield favorable effects than might be the case with a set of practices that was more representative of the program nationally.

## 2. Method

### 2.1. Sample

Two BBBSA agencies on the West Coast enrolled youth in the study. Each agency served a predominantly urban area. One agency ("A") had been providing mentoring services for 7 years at the time the study began and had served 1,323 youth in its CBM program in the year prior to the study's initiation; the other agency ("B") had been providing mentoring services for 48 years and served 1,157 youth in its CBM program the prior year. All youth between the ages of 9 and 14 whose parent or guardian applied for them to participate in the agency's CBM program and who were approved for program participation following the agency's standard application process were eligible for the study and approached about participating (see Procedures below for details). The

enrollment period lasted about a year, beginning in Spring, 2008 and ending in Spring, 2009. A total of 764 youth were enrolled in the study. Each of these youth was randomly assigned by researchers with equal probability either to be immediately eligible for being matched with a mentor (treatment group;  $n = 379$ ) or to remain on the program waitlist for 13 months (control group;  $n = 385$ ), after which they could be matched with a mentor. A total of 199 of these youth were enrolled in the study with one or more siblings. In these cases, siblings were randomly assigned together, as one unit, receiving the same assignment.

Follow-up youth and/or parent surveys were obtained for 85.6% of the sample ( $n = 654$  for parent and/or youth;  $n = 629$  for youth;  $n = 634$  for parents), with comparable rates for treatment (87.0%) and control (84.2%) groups. This sample of 654 youth with available data on the 13-month outcomes constitutes the analytic sample for the present research (see discussion of missing data under Data Analysis below).<sup>1</sup>

### 2.2. Procedures

Families were recruited for the study as part of program enrollment. Program staff explained to each parent/guardian and youth that study participants who were randomly assigned to the treatment group would be prioritized for matching with a mentor, and that if the parent chose not to participate, in all likelihood, his or her child would not be matched with a mentor until the study was completed. Study participation required written parent/guardian consent and youth assent which were obtained for all but approximately 1% of recruited families. Following consent, an agency staff person administered the baseline survey to the youth in the agency's office (or at the youth's home in a quiet, private space) while the youth's parent completed their baseline survey in another room. Youth completed the survey online as staff read each question out loud. The survey session took about 30 min. To ensure that youth assigned to the treatment group were matched as soon as possible after their baseline assessment, often youth and parents did not complete their surveys until a potential mentor had been located. After baseline survey completion, the youth was randomly assigned to condition, and those assigned to the treatment group were matched with a mentor as soon as was feasible. Random assignment was carried out by a survey research firm contracted to assist with the study. Thirteen months after study enrollment, surveys were administered a second time to the youth and parent/guardian. Parents completed the survey independently, online, while youth completed the survey, administered by the research firm, over the phone. Prior to survey administration, youth were mailed or emailed an "answer card" which listed all eight response sets used in the survey, each labeled with a different letter (A through H) and printed in different style and color fonts. During administration, when response sets changed, youth were directed to refer to the relevant response set in their answer card. Youth received a \$10 gift card for completing this survey.

Program implementation data were submitted by each agency on a quarterly basis to an intermediary organization that was overseeing programmatic implementation for a larger project in which the agencies were participating. All study procedures were approved by Public/Private Ventures' Institutional Review Board.

<sup>1</sup> To further investigate attrition, analyses were conducted comparing those youth with and without 13-month survey data on demographic variables, agency, and scores on outcome measures at study baseline. Results revealed three statistically significant ( $p < .05$ ) differences. Compared to youth with follow-up data, those without follow-up data ( $n = 110$ ) were less likely to be from single-parent households (35.9% vs 53.1%), reported higher levels of depressive symptoms (Cohen's  $d = 0.21$ ), and were more likely to report having skipped school (25.9% vs 11.2%).

### 2.3. Program

As noted, the randomized controlled trial involving the two BBBSA agencies was conducted as part of a larger project (see [Herrera et al., 2013](#) for details). As part of this initiative, the two agencies agreed to adopt practices within their CBM programs that went beyond BBBSA Standards of Practice in effect at the time of the study ([BBBSA, 2003](#)). First, each participating agency agreed to provide (at a minimum) a two-hour group, in-person, curriculum-based training to mentors either before or within one month of the start date of their match. The training needed to include active, experiential learning. BBBSA standards at that time required only that an opportunity for training be provided to each mentor, and neither agency had provided training of the type and timing required for the initiative. For example, Agency A had required a one-hour one-on-one training/orientation, while Agency B had required a one-on-one orientation and offered a two-hour group training, but the latter was not required. Second, the agencies agreed to contact the mentor, youth, and parent/guardian in each match via phone or in person every month for monitoring and support. BBBSA standards required monthly support contacts during the first year of the match, but allowed contacts with the family to alternate between the youth and parent, and allowed contact to occur via email. Finally, participating agencies also agreed to establish the expectation that each match meet at least three times a month, with each contact lasting at least two hours. When a match could not meet in person or via phone, reciprocated email communication could serve as a contact once every three months. Mentors and youth/parents also were asked to commit to staying in the mentoring relationship at least 18 months. BBBSA standards at that time did not address frequency of mentor-youth contact or commitment duration of the match parties. However, per common practice, both agencies prior to the study had an expectation of at least two contacts per month (Agency A expected at least three) and a commitment duration of at least one year.

As is typically the case in BBBSA CBM programs, youth were almost exclusively matched with same-gender mentors, with the exception being 13 male youth (ranging in age from 9 to 13 years old) who were paired with female mentors. Mentors were, on average, 33.40 years old ( $SD = 10.93$  years), with 3.6% African American, 2.5% American Indian or Alaska Native, 7.2% Asian American, 2.9% Latino/a, 1.8% Multi-racial or another identified race/ethnicity, and 81.9% White. About half (51.3%) of the mentors had previously worked with youth in a volunteer setting, 40.4% had professional experience working with youth (e.g., teacher, youth worker), 22.4% had previous experience mentoring youth in BBBS or another program, and 14.8% had children.

### 2.4. Measures

#### 2.4.1. Program implementation and delivery

Program implementation data reported by each agency included information for each mentoring relationship on the mentor's training attendance (i.e., whether and when the mentor attended training and the duration of the training); when agency staff had support contacts with the mentor, as well as the youth and their parent or other primary caregiver; and dosage (i.e., monthly data on the frequency of mentor-youth meetings and their duration as well as the date the mentoring relationship was established and the number of months it remained active). The intermediary summarized these data and shared it with each agency on a quarterly basis. Data at the individual match level were also shared for matches that did not meet expected thresholds (e.g., had not been contacted as frequently as outlined in practice expectations). For matches that did not meet the stated thresholds, the relevant agency was required to identify in writing proposed steps they would take to have those expectations met.

#### 2.4.2. Mentoring relationship quality

As part of the 13-month survey, youth who had been matched with a

mentor reported on their experiences in their mentoring relationship. If youth had been matched with more than one mentor, they were asked to respond about the mentor with whom they had been matched most recently. Individual items asked youth to report their feelings of closeness to the mentor ("How close do you feel to your mentor?"; Not close at all (1), Not very close (2), Somewhat close (3), or Very close (4); [Herrera et al., 2007](#)) and their expectations for the duration of the relationship ("I think I will be friends with my mentor for a long time"; Not at all true (1), Not very true (2), Sort of true (3), Very true (4); item developed for this research).

Youth also completed several multi-item scales assessing youth-centeredness (5 items, e.g., "My mentor almost always asks me what I want to do",  $\alpha = 0.80$ ; adapted from [Grossman & Johnson, 1999](#)), growth/goal focus (5 items, e.g., "My mentor helps me to set and reach goals",  $\alpha = 0.81$ ; [DuBois & Keller, 2017](#)), help with coping (3 items, e.g., "When something is bugging me, my mentor listens while I get it off my chest",  $\alpha = 0.66$ ; [Rhodes et al., 2005](#)), and negative emotional experiences (6 items, e.g., "When my mentor gives me advice, she/he makes me feel kind of stupid",  $\alpha = 0.66$ ; [Rhodes et al., 2005](#)). All items on these scales were rated on a 4-point Likert scale: Not at all true (1), Not very true (2), Sort of true (3), and Very true (4).

For ease of presentation, cutoff scores were used to group mean scores into 3 categories representing relatively high, moderate, and low levels of endorsement of the items comprising each scale. For youth centeredness, growth/goal focus, and help with coping, mean scores of 3.50 or higher were categorized as high, 2.50 through 3.49 were categorized as moderate, and scores below 2.5, the conceptual midpoint of the response scale, were categorized as low. Although normative data were not available to inform these cut-offs, the decision to require a high level of endorsement is consistent with the widely observed tendency for scores on youth-report measures of positive aspects of their mentoring relationships in BBBS and other similar programs to be concentrated toward the high end of the possible distribution of scores (i.e., negatively skewed; see, e.g., [Ferro et al., 2014](#)). Similarly, because negative emotional experiences tend to be reported relatively infrequently, for this scale, a score of 1.5 or lower was required for categorization as low, with scores 1.51 through 2.50 categorized as moderate and greater than 2.50 as high.

#### 2.4.3. Covariates

Several demographic variables, as reported by the youth's parent, were included as covariates in analyses of program impact.

**2.4.3.1. Gender, age and race/ethnicity.** Parents responded to three questions asking about the child's gender (male or female), date of birth, and race/ethnicity (i.e., African American, Black; American Indian or Alaska Native; Asian; Caucasian, White; Latino, Hispanic; Pacific Islander; Other) to which parents could check all that applied. Date of birth was subtracted from the date of the baseline survey to calculate the child's age.

**2.4.3.2. Receipt of free/reduced-price lunch.** Parents were asked "Does your child receive free or reduced-priced lunch at school?" and could reply, No, Yes, or I don't know.

**2.4.3.3. Parent education level.** Parents were asked, "What is the highest grade you completed?" and could select from six options: 8th grade or under, 9th–11th grade, 12th grade, 1–2 years of college, 3–4 years of college, and more than 4 years of college.

**2.4.3.4. Single-parent household.** To determine single-parent status, parents were asked: "Think about where your child lives. Who lives there with him or her?" Parents could check all that apply of the following 12 categories: Mother, Father, Brother/stepbrother(s), Sister/stepsisster(s), Stepmother, Stepfather, Grandmother, Grandfather, Aunt,



Uncle, Foster Parents, and Other. Parents were classified as “single parent” if they either: (1) checked only one category except when foster parents was selected because the category refers to multiple adults; or (2) checked two or more categories, but either of the sibling categories were selected as the only other category(ies).

#### 2.4.4. Youth outcomes

Youth outcomes were assessed using the pre-specified primary outcome measures from the broader initiative as well as parent-reported measures where available. Unless otherwise noted, each measure was completed by youth and was scored by taking the mean of the items.

**2.4.4.1. Depressive symptoms.** Depressive symptoms were assessed using the well-validated Short Mood and Feelings Questionnaire (SMFQ; Angold et al., 1995), which consists of 13 items. Each item is rated by the youth with respect to mood or feelings during the past two weeks as Not true (1), Sometimes true (2), or True most of the time (3). Examples include, “I felt miserable or unhappy” and “I thought nobody really loved me.” Internal consistency was satisfactory at both baseline and the 13-month assessment ( $\alpha = 0.89$  and  $0.87$ , respectively).

**2.4.4.2. Emotional symptoms.** The Emotional Symptoms subscale of the well-validated Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) was completed by the youth’s parent/guardian. The items on the SDQ ask about how true each statement is for the child over the last six months. Response options for this and all other SDQ subscales are Not true (0), Somewhat true (1), and Certainly true (2), and responses for each subscale are summed. Sample items for the 5-item Emotional Symptoms subscale include, “Is often unhappy, depressed or tearful” and “Often complains of headaches, stomachaches or sickness.” Internal consistency was satisfactory ( $0.73$  and  $0.71$  at baseline and 13-month assessments, respectively).

**2.4.4.3. Prosocial behavior (youth report).** To assess prosocial behavior, youth completed the Prosocial Behavior subscale of the Social-Emotional and Character Development Scale (SECD; Ji et al., 2013). This subscale consists of 6 items, each rated on a 4-point scale from Not at all true (1) to Very true (4). Sample items include, “I treat my friends the way I like to be treated” and “I am nice to kids who are different from me.” Prior research has supported the reliability and validity of this measure (Ji et al., 2013), including sensitivity to effects of a school-based positive youth development intervention (Lewis et al., 2016). Internal consistency was satisfactory at baseline ( $0.75$ ) and somewhat lower than desirable at 13 months ( $0.64$ ).

**2.4.4.4. Prosocial behavior (parent report).** Parents completed the 5-item Prosocial subscale of the SDQ (Goodman, 1997). Sample items include, “Is considerate of other people’s feelings” and “Is helpful if someone is hurt, upset or feeling ill.” Internal consistency was  $0.74$  at baseline and  $0.61$  at 13 months.

**2.4.4.5. Social acceptance.** Social acceptance was assessed using an adaptation of items comprising the 6-item Social Competence subscale of the Self-Perception Profile for Children (SPPC; Harter, 1985). Responding to items on the SPPC requires a two-step process in which youth choose which of two alternative statements is most like them (e.g., “Some kids have a lot of friends BUT Other kids wish they had more friends”) and then rate the extent to which that statement is true of them. In addition to questions in this form requiring more time and effort than a more typical single-statement format, experimental research with adolescents suggests that the two-statement format results in decreased measurement accuracy (Yeager & Krosnick, 2011). For these reasons, the scale items were simplified for the present study by selecting one of the two statements for each item and having youth rate how true this statement was of them on the same 4-point scale described

above for the SECD. Sample items include, “I have a lot of friends” and “I wish that more people my age liked me” (reversed). Internal consistency was satisfactory ( $0.76$  at both baseline and 13 months).

**2.4.4.6. Peer problems.** Parents completed the 5-item Peer Problems subscale of the SDQ (Goodman, 1997). Sample items include, “Has at least one good friend” (reversed) and “Often fights with other youth or bullies them.” Internal consistency was  $0.60$  at baseline and  $0.58$  at 13 months.

**2.4.4.7. Parent-child relationship quality.** Quality of the parent-child relationship was assessed using a 7-item measure adapted from the Trust scale of the Inventory of Parent and Peer Attachment (Armsden & Greenberg, 1987). Sample items include, “I trust my parent” and “My parent accepts me as I am.” Each item was rated on a 4-point scale from Hardly ever (1) to Pretty often (4). Internal consistency was satisfactory ( $0.85$  and  $0.83$  at baseline and 13 months).

**2.4.4.8. Misconduct (youth report).** Youth completed a 5-item scale of misconduct (Brown et al., 1986, as adapted by Posner & Vandell, 1994). Each item asked how often the youth had engaged in the behavior described in the past three months. Examples include, “Taken something on purpose that didn’t belong to you” and “Hit someone because you didn’t like something they said or did.” Response options were: Never in my life; I have done this, but not in the last 3 months; I did it 1–2 times in the last 3 months; and I did it 3 or more times in the last 3 months. Due to anticipated low frequency of endorsement, this measure was pre-specified to be scored as a dichotomous variable indicating whether the youth reported having engaged in any of the behaviors asked about over the preceding 3 months.

**2.4.4.9. Conduct problems (parent report).** Parents completed the 5-item Conduct Problems subscale of the SDQ (Goodman, 1997). Sample items include, “Often lies or cheats” and “Often fights with other youth or bullies them.” Internal consistency was satisfactory ( $0.73$  at baseline and  $0.75$  at 13 months).

**2.4.4.10. Self-perception of academic ability.** Self-perception of academic ability was assessed using the six-item subscale from the Self-Perception Profile for Children (SPPC; Harter, 1985) with a Likert adaptation similar to that used for the 6-item Scholastic Competence subscale of the SPPC. Sample items include, “I often forget what I learn” (reversed) and “I am very good at my schoolwork.” Internal consistency was  $0.68$  at baseline and  $0.64$  at 13 months.

**2.4.4.11. Skipping school.** Following Herrera et al. (2007), youth were asked two questions about how often they had skipped school over the past three months: “Skipped school without your parent or other person who has raised you knowing” and “Skipped a class without being allowed.” Response options were the same as those for the youth-reported measure of misconduct described above. Given relatively low frequency of endorsement, this outcome was pre-specified to be represented as a dichotomous measure indicating whether either type of skipping was reported over the prior 3 months. Furthermore, following Herrera et al. (2007), it was decided in advance that impact analyses for this outcome would be limited to those youth who did not report skipping school at study baseline, thus allowing for examination of program effects on onset of skipping school.

**2.4.4.12. Academic performance.** Youth were asked to report on the grades or marks they were currently receiving in school in each of four subject areas (i.e., math, reading or language arts, social studies, science). There were five response options ranging from F (not very good at all) (1) to A (excellent) (5).

**2.4.4.13. Hyperactivity.** Parents completed the 5-item Hyperactivity subscale of the SDQ (Goodman, 1997). Sample items include, “Restless, overactive, cannot stay still for long” and “Easily distracted, concentration wanders.” Internal consistency was satisfactory (0.79 at baseline and 0.80 at 13 months).

**2.4.4.14. Total Difficulties Score on the SDQ.** The Total Difficulties Score on the SDQ, which is the sum of scores across subscales, excepting Prosocial, also was included as an outcome measure. Internal consistency was satisfactory (0.83 at baseline and 0.86 at 13 months).

**2.4.4.15. Composite measures.** To facilitate obtaining an estimate of the overall effect of program participation across outcomes, we created composite measures based on the average of each outcome measure after first standardizing each measure by z-scoring; negatively valenced measures were inverted so that higher scores represented more favorable outcomes in all cases. One outcome, skipping school, was not included, because it was pertinent only to the subsample of youth who did not report having skipped school at baseline rather than the full sample as with all of the other outcomes. Three separate composites were created for all remaining outcomes to explore potential differences based on informant. These composites consisted of: (1) both youth- and parent-reported outcomes; (2) only youth-reported outcomes; and (3) only parent-reported outcomes. Parallel versions of each composite were computed based on 13-month and baseline scores, respectively.

## 2.5. Data analysis

Preliminary analyses provided descriptive statistics on study control and outcome measures at baseline and tested their equivalence across the treatment and control groups. Descriptive statistics on program implementation measures and youth reports of the quality of their mentoring relationship also were estimated.

Primary intent-to-treat analyses then estimated effects of CBM program participation (i.e., assignment to the treatment group) on each of the youth outcomes. Continuous outcomes were analyzed using ordinary least squares regressions, while the two dichotomous outcomes (youth-reported misconduct and skipping school) were analyzed using logistic regressions. Program effects on each composite were examined using the same regression approach used for individual outcome measures. For example, for the regression examining effects on the composite measure based on only youth-reported outcome measures, an equivalently computed composite based on the same youth-reported outcomes at baseline was included as a control.

Along with a treatment dummy (treatment = 1, control = 0), each regression included the following control variables: baseline score on the outcome, agency, youth gender, ethnicity (African American, Hispanic, White, other), age, receipt of free/reduced-price lunch (yes/no, with “don’t know” responses treated as missing), parent education level, and single-parent household. Because the treatment and control groups showed non-significant trends toward differing on two of the outcome measures at baseline, peer problems and skipping school (see Table 1), these variables were included as additional control measures in all analyses. As a sensitivity analysis, analyses were also conducted accounting for the non-independence of observations of siblings within families. To accomplish this, mixed models were fit in which family was included as a random intercept. The intraclass correlations associated with family membership for outcome measures are reported as well to provide information on the extent of dependency that existed among siblings.

When data were missing on an outcome measure at baseline, which was rarely the case (4.3% of youth), we substituted the mean value of the remainder of the sample and also included a dummy variable in regression analyses indicating whether the variable was missing for each participant (Deke & Puma, 2013). Otherwise, complete case analysis was used, such that youth with missing data on the outcome measure at

**Table 1**

Descriptive Statistics for Youth Demographic Characteristics and Outcome Measures at Study Baseline

Demographic Characteristic or Outcome Measure	Full Sample	Control Group	Treatment Group	Test for Group Difference (two-tailed p-value)
Agency (% Agency A)	58.6	59.1	58.0	0.77
Age (years)	11.41 (1.64)	11.38 (1.57)	11.43 (1.70)	0.67
Gender (% female)	42.2	41.8	42.1	0.83
Race/ethnicity (%)				0.07
African American	27.7	31.3	23.8	
Hispanic	21.4	18.2	24.8	
White	40.2	39.1	41.4	
Other	11.3	10.0	10.7	
Free/reduced-price lunch (%)	75.1	78.2	71.9	0.07
Parent education level	3.60 (1.35)	3.54 (1.32)	3.67 (1.38)	0.43
Single-parent household (%)	53.1	54.6	51.6	0.21
Depressive symptoms (youth-report)	1.55 (0.44)	1.55 (0.45)	1.55 (0.44)	0.94
Emotional symptoms (parent-report)	3.71 (2.52)	3.59 (2.48)	3.84 (2.55)	0.21
Prosocial behavior (youth-report)	3.51 (0.45)	3.51 (0.44)	3.51 (0.45)	0.92
Prosocial behavior (parent-report)	7.57 (2.11)	7.61 (2.04)	7.53 (2.18)	0.69
Social acceptance (youth-report)	2.66 (0.71)	2.66 (0.69)	2.66 (0.73)	0.96
Peer problems (parent-report)	2.92 (2.13)	2.81 (2.14)	3.06 (2.12)	0.13
Parent-child relationship quality (youth-report)	3.34 (0.63)	3.35 (0.63)	3.33 (0.63)	0.59
Hyperactivity (parent-report)	4.98 (2.76)	4.93 (2.69)	5.04 (2.84)	0.61
Misconduct (youth-report; %)	47.5	48.1	47.0	0.79
Conduct problems (parent-report)	2.59 (2.19)	2.54 (2.25)	2.64 (2.14)	0.56
Self-perception of academic ability (youth-report)	2.88 (0.60)	2.86 (0.59)	2.89 (0.62)	0.55
Skipping school (youth-report; %)	10.6	8.4	12.9	0.06
Academic performance (youth-report)	3.57 (0.89)	3.56 (0.85)	3.58 (0.89)	0.81
Total Difficulties Score on SDQ (parent-report)	14.20 (6.88)	13.82 (6.91)	14.59 (6.85)	0.16

Notes. N for full sample ranges from 643 to 654. For continuous variables, means are provided with standard deviation in parentheses. Group differences were tested using an independent groups *t*-test for continuous variables and using the chi-square test for categorical variables. SDQ = Strengths and Difficulties Questionnaire.

follow-up were excluded. Multiple imputation is a popular alternative for handling missing data. However, recent literature on randomized controlled trial methodology has noted advantages of complete case analysis (assuming control for any baseline measures on which treatment-control group differences are evident) relative to multiple imputation (Deke & Puma, 2013; Sullivan et al., 2018). These advantages include complete case analyses being less complex to implement (Deke & Puma, 2013) as well as the potential for multiple imputation to reduce statistical power due to variability in results across imputed datasets (Sullivan et al., 2018; Jakobsen et al., 2017). All analyses were carried out using Version 27 of the IBM SPSS Statistics software.

Effect sizes for continuous outcomes were computed by taking the unstandardized coefficient on the dummy variable for treatment status from the relevant regression analysis, which represented the estimated impact on the outcome at 13 months, and dividing this by the pooled

standard deviation of the 13-month outcome measure. The resulting effect size, thus, is in the form of Cohen's  $d$ , but with the difference between treatment and control groups on the outcome adjusted for baseline scores on the outcome and other control variables (Lipsey & Wilson, 2001). For the dichotomous outcomes, we converted the odds ratio for the group difference on the outcome to the approximate equivalent of Cohen's  $d$  using the Cox index (i.e., the natural log of the odds ratio divided by 1.65; see Chinn, 2000; Sánchez-Meca et al., 2003).

### 3. Results

#### 3.1. Descriptive statistics for study measures and tests of baseline equivalence

Descriptive statistics for study measures for the full sample as well as for the control and treatment groups are shown in Table 1 along with the results of tests of equivalence on control variables and outcome measures at baseline across treatment and control groups. Demographic characteristics of youth assigned to each group were largely comparable across treatment and control groups, although there were trends ( $p < .10$ ) toward significant differences for race/ethnicity and free/reduced-price lunch status. There was a larger proportion of youth in the control group who were African American (31.3%) than in the treatment group (23.8%) and, conversely, a smaller proportion of control group youth who were Latino/a (18.2%, compared to 24.8% in the treatment group). Control group youth were also somewhat more likely to receive free/reduced-price lunches (78.2%) than those in the treatment group (71.9%). For outcome measures, as noted previously, there were non-significant trends toward group differences for the measures of skipping school ( $p < .10$ ) and peer problems ( $p < .15$ ), with a larger proportion of youth in the treatment group reporting skipping (12.3%) compared to those in the control group (8.4%), and youth in the treatment group were rated somewhat higher in peer problems by their parents (Cohen's  $d = 0.12$ ).

#### 3.2. Program implementation and mentoring relationship quality

Nineteen youth in the treatment group were not able to be matched with a mentor. Of those matched, the match nearly always took place no more than one month after study enrollment and completion of the baseline surveys (89.7% were matched within one month;  $M = 0.63$  months). There also were 41 youth who were matched with one additional mentor during the study period after the initial match ended, and 2 youth who were matched 2 additional times. As noted, at 13 months, youth in the treatment group reported on their relationship with the mentor with whom they were currently or most recently matched. To maintain consistency with these data, the program implementation data (e.g., mentor training) are also reported for this most recent match.

##### 3.2.1. Mentor training

The mentors of slightly more than one-third of the treatment group youth (35.8%) attended program training before or within a month of their match start date; an additional 14.5% attended training after that window. The average training duration was slightly longer than 2 h ( $M = 2.16$ ,  $SD = 0.39$ ).

##### 3.2.2. Support

On average, mentors had support contacts (via phone or in person) with agency staff 67.4% of the months that the match was active during the study period. Youths' caregivers and youth themselves received support contacts in 67.4% and 60.0% of months, respectively.

##### 3.2.3. Mentoring dosage

At the time of the 13-month assessment, youth in the treatment group who had been matched with a mentor had, on average, received approximately 9 months of mentoring from the mentor with whom they

were matched at, or closest to, follow-up ( $M = 9.59$ ,  $SD = 4.18$ ), with about half (50.1%) having received at least 12 months of mentoring. Two-thirds of the youth who had been matched (66.8%) were in an active match with a mentor at the time of their 13-month assessment.

Youth in the treatment group met, on average, approximately two times per month ( $M = 2.14$ ,  $SD = 0.81$ ) for a total of about 6 h ( $M = 6.13$ ,  $SD = 2.62$ ). Only a small percentage of youth (15.0%) met with their mentor 3 or more times in at least 70 percent of the months they were matched. However, when lowering the threshold to 2 or more meetings, this percentage increased to approximately half (49.5%).

#### 3.2.4. Mentoring relationship quality

Two-thirds of youth (66.9%) reported feeling very close to their mentor, with the majority of the remaining youth reporting feeling somewhat close (23.2%) and only a small minority reporting feeling not very close (6.5%) or not close at all (3.4%). Similarly, a little over three-quarters of youth (76.1%) responded that it was "very true" that they expected to have an enduring friendship with their mentor (12.6%, 5.8%, and 5.5% for "sort of true," "not very true" and "not at all true," respectively).

With respect to the multi-item scales, using the predefined criteria described in the Method section, the overwhelming majority of youth (84.0%) reported a high level of youth-centeredness in their mentoring relationships (13.7% and 2.4% reported moderate and low levels, respectively). Similarly, three-fourths of youth (75.4%) reported a high level of help with coping from their mentors (18.8% and 5.8% for moderate and low levels, respectively). A smaller proportion of youth, about half (49.3%), reported a high level of growth orientation in their mentoring relationships (40.1% and 10.6%, for moderate and low levels, respectively). Finally, for the most part, youth reported only a low level of negative emotional experiences with their mentors (84.0%, 13.7%, and 2.4% for low, moderate, and high levels, respectively).

#### 3.3. Program effects

##### 3.3.1. Individual outcome measures

As can be seen in Table 2, intent-to-treat analyses revealed significant differences between treatment and control groups on only one of the eight youth-reported outcome measures: depressive symptoms ( $b = -0.057$ ,  $p < .05$ ,  $d = 0.146$ ). For the remaining youth-reported outcome measures, estimated effects were in directions favoring the treatment group for all but two measures: misconduct and skipping school.

As shown in Table 3, the parallel analyses for parent-reported outcome measures revealed statistically significant differences that favored the treatment group on four of six measures: emotional symptoms ( $b = -0.506$ ,  $p < .01$ ,  $d = 0.212$ ), peer problems ( $b = -0.500$ ,  $p < .001$ ,  $d = 0.253$ ), conduct problems ( $b = -0.302$ ,  $p < .05$ ,  $d = 0.138$ ), and SDQ Total Difficulties ( $b = -1.560$ ,  $p < .001$ ,  $d = 0.220$ ). There also were differences favoring the treatment group that approached statistical significance for parent ratings of prosocial behavior ( $b = 0.232$ ,  $p < .10$ ,  $d = 0.130$ ).

To summarize, estimated program effects were in directions favoring the treatment group for 12 of the 14 measures. These differences were statistically significant ( $p < .05$ ) or approached this threshold ( $p < .10$ ) for 6 of the 14 measures, 5 of which were parent-report. Estimated effect sizes for these measures ranged from 0.130 to 0.253 and thus were in the range of what would be considered differences of small magnitude by traditional rules of thumb for effect size interpretation (Cohen, 1984).

##### 3.3.2. Composite outcome measures

As shown in Table 4, analyses revealed a significant difference between treatment and control groups on the composite measure that included both youth- and parent-reported measures ( $b = 0.113$ ,  $p < .001$ ,  $d = 0.213$ ) and on the composite that was based on only parent-reported measures ( $b = 0.141$ ,  $p < .001$ ,  $d = 0.215$ ). Furthermore, the difference between treatment and control groups on the composite measure that

**Table 2**

Results of Intent-to-Treat Analyses of Effects of Big Brothers Big Sisters Community-Based Mentoring Program on Youth-Reported Outcomes

Predictor	Depressive Symptoms	Prosocial Behavior	Social Acceptance	Parent-Child Relationship Quality	Misconduct	Self-Perception of Academic Ability	Skipping School	Academic Performance
BBBS CBM program (vs control group)	−0.057* (0.028) <i>d</i> = 0.146	0.011 (0.025) <i>d</i> = 0.032	0.048 (0.046) <i>d</i> = 0.070	0.059 (0.038) <i>d</i> = 0.111	1.05 <i>d</i> = −0.030	0.041 (0.039) <i>d</i> = 0.073	1.23 <i>d</i> = −0.121	0.069 (0.057) <i>d</i> = 0.091
Baseline score on outcome	0.368*** (0.034)	0.253*** (0.029)	0.505*** (0.035)	0.318*** (0.032)	3.18***	0.474*** (0.033)	—	0.270*** (0.034)
Agency <sup>a</sup>	0.046 (0.030)	0.017 (0.026)	−0.021 (0.048)	−0.041 (0.040)	1.07	−0.020 (0.040)	0.50 <sup>†</sup>	−0.022 (0.059)
Youth age	0.003 (0.009)	−0.002 (0.008)	0.027 (0.015)	−0.056*** (0.012)	1.01	−0.016 (0.012)	1.61***	−0.051** (0.018)
Youth gender: Female <sup>b</sup>	−0.022 (0.029)	0.063* (0.026)	0.034 (0.046)	−0.022 (0.039)	1.06	0.058 (0.039)	0.99	0.104 <sup>†</sup> (0.058)
Race/ethnicity: African American <sup>c</sup>	−0.079 (0.052)	0.030 (0.047)	0.096 (0.084)	−0.020 (0.070)	1.04	0.068 (0.071)	0.54	−0.048 (0.105)
Race/ethnicity: Hispanic <sup>c</sup>	−0.076 (0.054)	0.034 (0.048)	0.126 (0.087)	−0.031 (0.072)	0.96	0.003 (0.073)	0.42	−0.035 (0.108)
Race/ethnicity: White <sup>c</sup>	−0.062 (0.050)	0.034 (0.045)	0.000 (0.080)	0.025 (0.067)	0.85	0.028 (0.068)	0.25	0.095 (0.100)
Free/reduced-price lunch	−0.017 (0.035)	−0.005 (0.031)	−0.035 (0.057)	0.035 (0.047)	1.16	0.031 (0.048)	1.23	−0.037 (0.071)
Parent education level	−0.002 (0.011)	−0.001 (0.010)	0.003 (0.018)	0.005 (0.015)	0.99	0.022 (0.015)	0.85	0.026 (0.023)
Single-parent household	0.017 (0.029)	−0.027 (0.026)	0.013 (0.046)	−0.004 (0.039)	0.92	0.031 (0.039)	1.59	0.015 (0.058)
Peer problems	0.015* (0.007)	−0.015* (0.006)	−0.025* (0.012)	−0.017 <sup>†</sup> (0.009)	1.07 <sup>†</sup>	−0.017 <sup>†</sup> (0.009)	1.20*	−0.017 (0.014)
Skipping school	0.027 (0.048)	−0.066 (0.042)	−0.017 (0.075)	−0.055 (0.064)	1.57	−0.081 (0.065)	—	−0.227* (0.094)

Notes. N = 628 or 629 with the exception of the analysis for Skipping School for which N = 557 due to exclusion of youth reporting skipping at baseline. Unstandardized regression coefficients with standard errors in parentheses are reported with the exception of the dichotomous outcomes, Misconduct and Skipping School, for which odds ratios are reported. *d* represents effect size calculated as described in text, with positive (negative) values indicating differences in the direction of more (less) favorable scores on the outcome measure for the treatment group relative to the control group. Dummy variables representing those with missing data on baseline measures of outcomes were included in models but are not shown (see text).

<sup>a</sup> Agency A = 0; Agency B = 1. <sup>b</sup> Male = 0; Female = 1. <sup>c</sup> Reference group is “Other”.

<sup>†</sup> *p* < .10; \* *p* < .05; \*\* *p* < .01; \*\*\* *p* < .001.

**Table 3**

Results of Intent-to-Treat Analyses of Effects of Big Brothers Big Sisters Community-Based Mentoring Program on Parent-Reported Outcomes

Predictor	Emotional Symptoms	Prosocial Behavior	Peer Problems	Hyperactivity	Conduct Problems	SDQ Total Difficulties
BBBS CBM program (vs control group)	−0.506** (0.161) <i>d</i> = 0.212	0.232 <sup>†</sup> (0.122) <i>d</i> = 0.130	−0.500*** (0.134) <i>d</i> = 0.253	−0.266 (0.170) <i>d</i> = 0.092	−0.302* (0.137) <i>d</i> = 0.138	−1.560*** (0.427) <i>d</i> = 0.220
Baseline score on outcome	0.446*** (0.035)	0.410*** (0.030)	0.492*** (0.033)	0.679*** (0.033)	0.613*** (0.033)	0.700*** (0.042)
Agency <sup>a</sup>	0.056 (0.168)	−0.025 (0.127)	−0.040 (0.140)	−0.257 (0.177)	−0.328* (0.143)	−0.105 (0.446)
Youth age	0.074 (0.050)	−0.092 (0.038)	0.021 (0.042)	0.087 (0.053)	0.081 <sup>†</sup> (0.043)	0.272* (0.133)
Youth gender: Female <sup>b</sup>	0.261 (0.164)	0.274* (0.124)	−0.082 (0.136)	−0.532** (0.175)	−0.036 (0.141)	−0.417 (0.435)
Race/ethnicity: African American <sup>c</sup>	0.020 (0.292)	0.288 (0.220)	−0.066 (0.243)	−0.261 (0.311)	0.083 (0.248)	0.003 (0.773)
Race/ethnicity: Hispanic <sup>c</sup>	0.183 (0.300)	0.356 (0.227)	−0.246 (0.251)	0.080 (0.321)	0.220 (0.256)	0.318 (0.798)
Race/ethnicity: White <sup>c</sup>	0.608* (0.278)	0.261 (0.211)	0.216 (0.233)	0.347 (0.298)	0.083 (0.237)	1.269 <sup>†</sup> (0.740)
Free/reduced-price lunch	−0.071 (0.201)	0.030 (0.152)	0.060 (0.168)	0.003 (0.213)	0.059 (0.172)	0.071 (0.535)
Parent education level	0.080 (0.064)	−0.137** (0.048)	0.015 (0.053)	0.036 (0.060)	0.131* (0.054)	0.238 (0.168)
Single-parent household	−0.139 (0.163)	0.091 (0.123)	−0.552*** (0.136)	−0.070 (0.172)	−0.125 (0.134)	−0.859* (0.433)
Peer problems	0.105* (0.042)	−0.043 (0.031)	—	0.002 (0.044)	0.013 (0.035)	−0.229 (0.138)
Skipping school	−0.023 (0.264)	−0.474* (0.200)	−0.071 (0.221)	0.141 (0.280)	−0.009 (0.229)	−0.118 (0.707)

Notes. N ranges from 631 to 634. Unstandardized regression coefficients with standard errors in parentheses are reported. SDQ = Strengths and Difficulties Questionnaire. *d* represents effect size calculated as described in text, with positive (negative) values indicating differences in the direction of more (less) favorable scores on the outcome measure for the treatment group relative to the control group. Dummy variables representing those with missing data on baseline measures of outcomes were included in models but are not shown (see text).

<sup>a</sup> Agency A = 0; Agency B = 1. <sup>b</sup> Male = 0; Female = 1. <sup>c</sup> Reference group is “Other”.

<sup>†</sup> *p* < .10; \* *p* < .05; \*\* *p* < .01; \*\*\* *p* < .001.

was based on only youth-reported measures approached significance (*b* = 0.067, *p* < .10, *d* = 0.115).

### 3.3.3. Sensitivity analyses

Intraclass correlations (ICC) based on family membership ranged from 0.00 (academic performance) to 0.37 (parent–child relationship quality) for the youth-reported outcome measures (average ICC = 0.15),

and from 0.00 (hyperactivity) to 0.29 (emotional symptoms) for the parent-reported outcome measures (average ICC = 0.13). ICCs for the composite measures were 0.16 for the overall composite, 0.13 for the composite based on youth-reported measures, and 0.16 for the composite based on parent-reported measures. When accounting for this non-independence in mixed model analyses, all statistically significant treatment–control differences on outcome measures and composites



**Table 4**

Results of Intent-to-Treat Analyses of Effects of Big Brothers Big Sisters Community-Based Mentoring Program on Composite Measures

Predictor	Youth- and Parent- Report Measures Composite	Youth-Report Measures Composite	Parent-Report Measures Composite
BBBS CBM program (vs control group)	0.113*** (0.032) <i>d</i> = 0.213	0.067 <sup>†</sup> (0.039) <i>d</i> = 0.115	0.141*** (0.039) <i>d</i> = 0.215
Baseline score on outcome	0.677*** (0.038)	0.538*** (0.036)	0.631*** (0.038)
Agency <sup>a</sup>	0.009 (0.033)	−0.035 (0.040)	0.045 (0.041)
Youth age	−0.017 <sup>†</sup> (0.010)	−0.016 (0.012)	−0.026* (0.012)
Youth gender: Female <sup>b</sup>	0.054 <sup>†</sup> (0.032)	0.048 (0.039)	0.039 (0.040)
Race/ethnicity: African American <sup>c</sup>	0.058 (0.059)	0.060 (0.071)	0.012 (0.071)
Race/ethnicity: Hispanic <sup>c</sup>	0.063 (0.060)	0.056 (0.073)	0.018 (0.074)
Race/ethnicity: White <sup>c</sup>	0.042 (0.056)	0.100 (0.068)	−0.057 (0.068)
Free/reduced- price lunch	0.019 (0.039)	0.011 (0.048)	−0.025 (0.049)
Parent education level	−0.005 (0.012)	0.010 (0.015)	−0.034* (0.016)
Single-parent household	0.027 (0.032)	−0.003 (0.039)	0.075 <sup>†</sup> (0.040)
Peer problems	−0.007 (0.009)	−0.028** (0.010)	0.016 (0.012)
Skipping school	−0.037 (0.054)	−0.072 (0.066)	−0.091 (0.065)

Notes. As described in the Method section, the three composites for youth-, parent-, and youth- and parent-report measures assess overlapping but not entirely parallel constructs. *N* ranges from 609 to 634. Unstandardized regression coefficients with standard errors in parentheses are reported. *d* represents effect size calculated as described in text, with positive (negative) values indicating differences in the direction of more (less) favorable scores on the outcome measure for the treatment group relative to the control group. Dummy variables representing those with missing data on baseline measures of outcomes were included in models but are not shown (see text).

<sup>a</sup> Agency A = 0; Agency B = 1. <sup>b</sup> Male = 0; Female = 1. <sup>c</sup> Reference group is “Other”.

<sup>†</sup> *p* < .10; \**p* < .05; \*\**p* < .01; \*\*\**p* < .001.

remained significant (*p* < .05) and all treatment–control group differences that had approached statistical significance (*p* < .10) continued to do so.

#### 4. Discussion

The goal of this study was to rigorously examine the effects of the BBBSA CBM program in two of its large agencies. The study was designed to strengthen the existing knowledge base by more fully examining implementation of the BBBSA CBM program model, assessing program outcomes through both youth and parent report, and including outcomes and composites of those outcomes to reflect potential program benefits across a range of outcomes.

Positive program impacts that achieved or approached statistical significance were evident for several of the outcomes tested, including emotional and depressive symptoms, peer problems, prosocial behavior, and conduct problems. Impacts were not found for any of the three school-related outcomes tested (i.e., academic performance, self-perception of academic ability, and skipping school), hyperactivity, parent–child relationship quality, and social acceptance.

All but one of the positive impacts were in areas reported by parents as opposed to youth, and our composite measure was significant for the parent-reported measures, but only approached significance for the youth-reported measures. This pattern is similar to that reported recently in a randomized controlled trial of Friends of the Children—a long-term mentoring program that incorporates both school- and community-based components (Eddy et al., 2017). In that research,

significant impacts favoring the treatment group were found in caregiver ratings of positive school behavior and getting into trouble in school, but no effects were found in either of the youth-reported outcomes measured (i.e., deviant peers and antisocial behavior). One explanation could be that parents are more attuned to and discriminating in their assessments of youth functioning. Mentoring may also be most likely to affect aspects of youth functioning that are observed by parents (e.g., better behavior around adults). Yet, stronger parent-reported impacts were evident even for two constructs that were measured through both parent and youth reports (prosocial behavior and misconduct)—both of which approached (or, reached, in the case of misconduct) significance for the parent- but not the youth-reported measure. In these cases, the measures differ slightly in ways that could help explain these differences. In the case of misconduct, the youth-reported measure includes a wider range of behaviors, some of which are fairly uncommon, particularly among younger youth (e.g., carrying a weapon, being a member of a gang). In this young sample, parent reports may have provided a more sensitive measure of misconduct than the youth-reported measure. In the case of prosocial behavior, the youth-reported measure is focused on behavior toward other youth, while the parent measure also includes behavior directed toward others more broadly, which would include adults—and the latter may be more susceptible to change through the provision of an adult mentor (and perhaps particularly noticeable by parents). To date, parent-report measures have been included in very few large-scale impact studies of mentoring. However, their perspectives may provide unique insights into their children's attitudes and behaviors, and they may be more attuned to changes in areas that are stimulated through the support of mentoring relationships. Thus, including their views in mentoring studies can provide a valuable perspective on their child's growth.

Depressive symptoms was the only youth-reported outcome for which a significant impact of program participation was evident. This measure has some overlap with parent-reported emotional symptoms, for which a similar benefit of program involvement was apparent. These are noteworthy findings. Depressive symptoms in middle childhood have been found to predict substance use in adolescence (Prinstein & La Greca, 2009; Saraceno et al., 2012) as well as aggression, lower self-esteem, poorer functioning in social relationships and education, and less job success in young adulthood (Aronen & Soininen, 2000). Further, although we did not measure clinical depression per se, scores on the SMFQ (Angold et al., 1995) have been shown to discriminate older adolescents who meet the International Classification of Diseases criteria for depression (ICD-10; World Health Organization, 1992; Turner et al., 2014), and childhood depression is associated with a wide range of serious difficulties including suicidal behavior and increased risk for substance abuse and teen pregnancy (for reviews, see Malhotra & Das, 2007; Groenman et al., 2017; see also Cash & Bridge, 2009). These findings suggest that the benefits of involvement in BBBSA CBM could potentially “trickle down” over time to other key areas of youth development.

The indicated benefits of program participation for emotional well-being are perhaps not surprising. Youth may simply “feel better” with positive attention from a caring adult and opportunities to engage in enjoyable activities that, in themselves, may boost a child's mood. In addition, as noted, program-based youth mentoring shares features with behavioral activation, an evidence-supported treatment for depression in which patients identify and engage in activities they enjoy (Cuijpers et al., 2007). Mentoring provided through a program such as BBBS CBM may also involve processes similar to those involved in cognitive-behavioral treatment of internalizing difficulties, including modeling and reinforcement of realistic appraisals, problem-solving and adaptive coping (Kerr & King, 2014). Although limited, findings from RCTs of the BBBS community-based program as implemented in Ireland and Canada also suggest benefits for the emotional well-being of participating youth. In the Ireland RCT, youth assigned to the program reported higher levels of hope than controls across the two-year follow-up period (Dolan,

Brady, O'Regan, Russell et al., 2010); qualitative findings, furthermore, suggested that youth receiving mentoring were happier, calmer and more confident (Dolan, Brady, O'Regan, Brumovska et al., 2010). The Canadian RCT (DeWit et al., 2007) was a small-scale feasibility study ( $N = 71$ ) and thus under-powered. However, it also yielded promising findings in this area. The study assessed outcomes over a one-year period and reported trends in the direction of program benefits for child-reported measures of emotional problems (on the SDQ) and social anxiety, albeit not for measures of parent-reported emotional problems or child-reported depressive symptoms. Future studies should explore mechanisms through which mentoring may improve youth's emotional well-being and examine possible impacts on other related outcomes such as loneliness and suicidal ideation. Programs should also consider highlighting to mentors their potentially important role in contributing to youth's emotional well-being.

Also of note are the study findings indicating benefits of program participation using composite measures that combined scores across outcome measures for child and/or parent reporters. This aspect of our results is in line with the orientation of the BBBS CBM program toward promoting gains across multiple areas of youth functioning in ways that foster overall trajectories of positive development. Effect sizes for each of the composite measures were larger than the average effect size of the corresponding individual measures of which they were comprised (i.e., for all measures: 0.213 vs 0.115; for youth-reported measures: 0.115 vs 0.079; for parent-reported measures: 0.215 vs 0.165). Similarly, the effect size for the Total Difficulties Score on the SDQ was larger than the average effect size of the four SDQ scales contributing to it (i.e., 0.220 vs 0.173). Statistically, these differences are in line with the potential for effect sizes for composites to exceed those for the individual measures from which they are derived. This is most likely to occur when the measures involved are not highly correlated and effect estimates are predominantly in the same direction across measures, as was largely the case in this study.

Conceptually, the relatively stronger effects for composite measures are consistent with a potential for youth to receive a cumulative benefit from mentoring received through BBBS CBM, across areas such as social competence, school performance, and problem behavior involvement. Benefits in any one area, however, may tend to be of too small a magnitude to reach a threshold of statistical significance, as was the case for all but one of the youth-report measures in the current study despite a fairly consistent directional pattern favoring the treatment group. Small average benefits in any given area across program participants may reflect the non-targeted nature of the BBBSA CBM program, which focuses broadly on relationship development, rather than on any particular outcome. A recent meta-analysis by Christensen et al. (2020), in fact, found that targeted mentoring programs yielded substantially larger effect sizes on the outcomes they targeted than did non-targeted programs (typically tested on a broader set of outcomes). Perhaps one of the strengths of non-targeted programs like BBBSA CBM is their ability to yield "small" benefits, but across a wide range of outcomes, whereas targeted programs may yield larger benefits, but only in the specific areas targeted by the programs. Prior studies have examined effects of BBBS CBM and other non-targeted mentoring programs with a similar orientation toward promoting outcomes across a broad range of areas, only with respect to individual measures. Consequently, their results may be underestimating benefits for participating youth when gauged from a more holistic developmental standpoint. Future studies should continue to explore how to assess impacts in ways that reflect the primary aims of programs, including their potential to improve youth's lives across a number of domains.

The current study found few significant impacts in the outcome domains it has in common with the earlier P/PV study of BBBSA CBM (Tierney et al., 1995). Several factors could account for these differences. First, by our 13-month follow-up, only about half of the treatment youth had received at least 12 months of mentoring, and the average match length was less than 10 months. The follow-up period for the

original P/PV study (Tierney et al., 1995) was 18 months, at which point, matched youth had met with their mentors an average of 12 months (Grossman & Rhodes, 2002). In a follow-up study using the P/PV sample, Grossman and Rhodes (2002) found that youth with 12 months of mentoring received more benefits than those with 6 to 12 months of mentoring. For example, effects on substance use and social acceptance achieved statistical significance only for those who had been mentored at least 12 months. Second, the P/PV trial was based on a substantially larger sample and thus was in a better position to detect small effects on program outcomes. Third, while youth in the treatment group in the current study reported fairly strong relationships with their mentors, only 15 percent met with their mentor 3 or more times in most of the months they were matched. In contrast, case managers reported that over 70 percent of youth in the original P/PV study met with their mentor at least three times a month, and close to half (45%) met at least once a week (Tierney et al., 1995). Because more consistent and frequent mentor-youth contact is associated with greater youth benefits (DuBois et al., 2002; Karcher, 2005), differences in meeting frequency across the two studies could partially explain their differences in impacts.

Implementation of program practices was variable across the matches involved in this study. This was especially true for initial mentor training (beyond basic program orientation), which was received by only about one in three mentors. Implementation of monthly support contacts with mentors and families was also inconsistent, with support contacts not taking place in at least one-third of the months that matches were active for any of the parties with whom they were expected to occur (i.e., mentors, youth, and parents). Research suggests that these variations may have affected program impacts. A recent meta-analysis (Goense et al., 2016), for example, of 17 studies of evidence-based interventions for youth with antisocial behavior found medium to large effects when treatment integrity was high and no significant effects when integrity was low.

Uneven implementation could be explained, in part, by the fact that aspects of these practices were new expectations associated with the initiative in which the agencies were participating. These expectations also differed from those for other matches on staff's caseloads that were not participating in the study (e.g., those that started the program prior to the study). As noted, at the time the study was conducted, BBBSA national standards required less frequent contact for youth and parents, and did not require mentor training, beyond a basic one-on-one orientation. Other studies of BBBSA mentoring reveal similarly low uptake of "enhanced" training opportunities by mentors (Courser et al., 2014; Peaslee & Teye, 2015). In a study including both BBBSA and other types of mentoring programs, Jarjoura et al. (2018) also noted inconsistency in implementation of new or enhanced practices and suggested a range of factors that might contribute to this, including organizational capacity and staff buy-in.

Yet, inconsistency in implementation was also evident for more established practices, such as monthly contacts with mentors during the first year of the match. Contributing factors may include difficulty reaching mentors, the mentors' understanding of the purpose of these contacts, and lack of a clear agreement before the matches were made that these contacts were a required program component. Broader agency factors may also have played a role. For example, one of the two agencies experienced economic and staffing challenges over the course of the study, which may have affected the support provided to study participants. Additionally, these practices are not only program dependent, but also very dependent on the actions of mentors and families—if the mentor or family fails to return staff's calls, not much can be done short of terminating involvement with the program and thus jeopardizing the well-being of the youth involved. The same holds for mentor training—without requiring attendance as a prerequisite for program participation, attendance will be varied. Although all are potential explanations for variations in implementation, it is unclear to what extent, if any, these factors may have influenced the study's findings.

Studies suggest that receipt of training and support is linked with match duration and quality as well as mentor program satisfaction. For example, the larger project associated with the current study found that matches were shorter and rated as lower in quality by youth when mentors did not receive early mentor training and/or consistent support contacts (Herrera et al., 2013). Keller et al. (2020), in a study of 55 youth mentoring programs, further found that mentors who reported no support contacts with program staff and those reporting very short contacts gave lower ratings of supervision quality, organizational culture, and service experience than mentors who reported longer contacts.

Despite strengths afforded by our randomized controlled trial design, limited attrition, and multi-informant reporting on outcomes, the study has several limitations. First, a lack of indicated impacts in this study may not reflect a true absence of effects. As noted, statistical power was limited for detecting relatively small effects on outcomes, such as those observed in the original RCT (Tierney et al., 1995), which had a larger sample. Future research should include RCTs with larger samples and consider meta-analysis as a method for aggregating results across well-controlled trials of the program for increased precision and sensitivity in impact estimation.

Second, the study did not observe outcomes over the full course of program involvement. Only half of study participants had received a year or more of mentoring, and two thirds of those who had been matched were still in active mentoring relationships at the time of our follow-up assessment, suggesting that the impacts we measured did not reflect full participation in the program for most of our sample and that a longer-term follow-up may have yielded different findings. Assessment of outcomes over longer periods of time will be needed to address this concern as well as the potential for effects to erode over longer intervals (see Herrera et al., 2007).

Third, the reports by treatment youth and their parents may have reflected some motivational bias. Parents of treatment youth may have reported that their child was doing well overall, due to cognitive dissonance (e.g., “My child must be doing better if I have put this much effort into program participation”; Festinger, 1957) or simply as a sign of appreciation or loyalty to the program. At the same time, control parents may have reported that their child was struggling overall, either to express discontent with not having been matched sooner or to stress that their child could still benefit from program participation. Future research should include additional sources of data, such as teacher reports and archival records of outcomes such as grades, school attendance, and delinquent offending.

On a related note, the decision to have program staff administer the baseline survey at enrollment (but not follow-up) reflects efforts to ensure that we collected surveys from as many families as possible at both time points. Although parents completed all surveys independently, youth completed surveys under different circumstances at the two time points (i.e., at baseline, in person, administered by program staff; at follow-up, by phone, administered by a survey firm). Because youth in both the treatment and control groups completed surveys using the same methodology at both baseline and follow-up, any differences in mode of data collection between baseline and follow-up would likely be experienced by both groups in the same way. However, any potential motivational differences (as discussed above) may have interacted with administration mode at follow-up in ways that could have affected findings (e.g., treatments being particularly likely to highlight “doing well” when completing the survey by phone).

Fourth, the study is also limited in its external validity. Both of the agencies were from the West Coast and are not representative of the full range of BBBSA programs. Furthermore, participating agencies were asked to implement practices beyond those required of BBBSA agencies at the time. This could have affected impacts in both positive and negative ways. For example, mentors may have received more supports than they normally would have which could have improved match quality; but staff may have been more burdened with providing these supports which could have had negative effects on support and

ultimately match quality. Research is needed with a larger number of agencies that are more representative of the BBBSA network in such factors as size, geographic location, age of program, and characteristics of the youth and families served. Similarly, the data for the study were collected more than 10 years ago. As is true of all programs, BBBSA CBM has continued to evolve over time, and the life challenges young people face are changing rapidly, making replication all the more important.

Finally, levels of internal reliability were only marginally acceptable for some of our measures, which suggests that measurement error could be a factor affecting our findings. This was true in particular for selected subscales of the Strengths and Difficulties Questionnaire, for which reliability estimates in a similar range also have been reported in prior research (see Stone et al., 2010, for a review). The reliability for parent-rated peer problems was particularly low, as was also the case in the review of Stone et al. (2010). The SDQ subscales reference relatively discrete behaviors, rather than attitudes. Thus, it could be argued that using internal consistency to gauge their reliability is not entirely appropriate (i.e., the items may be best conceptualized as causal/formative rather than reflective indicators of the constructs involved, such that their aggregates do not represent meaningful constructs or indices; for discussion, see Bollen & Lennox, 1991). Nevertheless, it will be important to replicate the present findings for the SDQ subscales, particularly for peer problems given that we did not find statistically significant impacts in youth-rated social acceptance. Likewise, internal consistency reliability for self-perception of academic ability was somewhat low and thus could have reduced sensitivity to detecting impacts on this outcome. Our approach to constructing composite measures of outcomes, although in line with other recent research (Resnjanskij et al., 2021), is also not without potential limitations, such as concerns that could be raised regarding the range of qualitatively distinct types of outcomes that were incorporated in these indices.

These limitations notwithstanding, the results of the present research add significantly to the evidence of the effectiveness of the BBBSA CBM program and, along with methodological features of the research (e.g., controlling for sibling dependencies, considering outcomes in the aggregate, comparing parent- versus child-reported impacts), provide a useful foundation for future research examining the impact of this and other similar mentoring programs. Especially important will be efforts to replicate the current findings in a larger, more geographically diverse sample of agencies implementing the BBBSA CBM program and to test whether program benefits are sustained over time.

#### Author note

Carla Herrera, David DuBois, and Jean Grossman serve in unpaid roles on the Research Advisory Council of Big Brothers Big Sisters of America (BBBSA). BBBSA did not have approval-granting authority for any aspect of this research. The opinions, findings, and conclusions expressed in this publication are those of the authors and do not necessarily reflect those of BBBSA.

The study was conducted while Carla Herrera and Jean Grossman were researchers with Public/Private Ventures.

#### Funding

This study was supported by funding to MENTOR Washington (formerly Washington State Mentors) and Public/Private Ventures from the Bill & Melinda Gates Foundation, Seattle, WA. The opinions, findings, and conclusions expressed in this publication are those of the authors and do not necessarily reflect those of the sponsor.

De-identified data that support the findings of this study are available from the corresponding author upon reasonable request, subject to approval of the funder (co-owner of the data).

#### CRedit authorship contribution statement

**Carla Herrera:** Conceptualization, Methodology, Investigation, Supervision, Project administration, Funding acquisition. **David L.**



**DuBois:** Conceptualization, Methodology, Formal analysis, Data curation. **Janet Heubach:** Conceptualization, Funding acquisition. **Jean B. Grossman:** Conceptualization, Methodology, Supervision, Funding acquisition.

### Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Carla Herrera, David DuBois, and Jean Grossman serve in unpaid roles on the Research Advisory Council of Big Brothers Big Sisters of America (BBBSA). BBBSA did not have approval-granting authority for any aspect of this research. The opinions, findings, and conclusions expressed in this publication are those of the authors and do not necessarily reflect those of BBBSA. This study was supported by funding to MENTOR Washington (formerly Washington State Mentors) and Public/Private Ventures from the Bill & Melinda Gates Foundation, Seattle, WA. The opinions, findings, and conclusions expressed in this publication are those of the authors and do not necessarily reflect those of the sponsor.

### Data availability

De-identified data that support the findings of this study are available from the corresponding author upon reasonable request, subject to approval of the funder (co-owner of the data).

### References

- Angold, A., Costello, E. J., Messer, S. C., Pickles, A., Winder, F., & Silver, D. (1995). Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *International Journal of Methods in Psychiatric Research*, 5(4), 237–249.
- Armsden, G. C., & Greenberg, M. T. (1987). The Inventory of Parent and Peer Attachment: Individual differences and their relationship to psychological well-being in adolescence. *Journal of Youth and Adolescence*, 16(5), 427–454. <https://doi.org/10.1007/BF02202939>
- Aronen, E. T., & Soininen, M. (2000). Childhood depressive symptoms predict psychiatric problems in young adults. *The Canadian Journal of Psychiatry*, 45(5), 465–470. <https://doi.org/10.1177/070674370004500507>
- Axford, N., Bjornstad, G., Matthews, J., Whybra, L., Berry, V., Ukoumunne, O. C., ... Warner, G. (2021). The effectiveness of a community-based mentoring program for children aged 5–11 years: Results from a randomized controlled trial. *Prevention Science*, 22(1), 100–112. <https://doi.org/10.1007/s11121-020-01132-4>
- Big Brothers Big Sisters of America (October, 2003). *Standards of practice for one-to-one service for programs operating within affiliated sponsoring organizations*. Big Brothers Big Sisters of America National Office.
- Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110(2), 305–314. <https://doi.org/10.1037/0033-2909.110.2.305>
- Brown, B. B., Clasen, D. R., & Eichler, S. A. (1986). Perceptions of peer pressure, peer conformity dispositions, and self-reported behavior among adolescents. *Developmental Psychology*, 22(4), 521–530. <https://doi.org/10.1037/0012-1649.22.4.521>
- Cash, S. J., & Bridge, J. A. (2009). Epidemiology of youth suicide and suicidal behavior. *Current Opinions in Pediatrics*, 21(5), 613–619. <https://doi.org/10.1097/MOP.0b013e32833063e1>
- Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, 19(22), 3127–3131. [https://doi.org/10.1002/1097-0258\(20001130\)19:22<3127::aid-sim784>3.0.co;2-m](https://doi.org/10.1002/1097-0258(20001130)19:22<3127::aid-sim784>3.0.co;2-m)
- Christensen, K. M., Hagler, M. A., Stams, G. J., 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, P. A. (1984). *Statistical power analysis for the behavioral sciences*. Erlbaum.
- Courser, M., Shamblen, S., Thompson, K. T., Young, L., Hamilton-Nance, S., Hutchins, M., & Wilbon, M. (2014). *Improving relationship outcomes using additional training and enhanced match support for mentors, final report*. Department of Justice: U. S. <https://www.ojp.gov/pdffiles1/ojdp/grants/251213.pdf>
- Cuijpers, P., Van Straten, A., & Warmerdam, L. (2007). Behavioral activation treatments of depression: A meta-analysis. *Clinical Psychology Review*, 27(3), 318–326. <https://doi.org/10.1016/j.cpr.2006.11.001>
- Claro, A., & Perelmiter, T. (2021). The effects of mentoring programs on emotional well-being in youth: A meta-analysis. *Contemporary School Psychology*, 1–13. <https://doi.org/10.1007/s40688-021-00377-2>. Advance online publication.
- Deke, J., & Puma, M. (2013). *Coping with missing data in randomized controlled trials*. Evaluation Technical Assistance Brief for OAH & ACYF Teenage Pregnancy Prevention Grantees, Brief 3.
- DeWit, D. J., Lipman, E., Manzano-Munguia, M., Bisanz, J., Graham, K., Offord, D. R., ... Shaver, K. (2007). Feasibility of a randomized controlled trial for evaluating the effectiveness of the Big Brothers Big Sisters community match program at the national level. *Children and Youth Services Review*, 29(3), 383–404. <https://doi.org/10.1016/j.chidyouth.2006.09.003>
- Dolan, P., Brady, B., O'Regan, C., Brumovska, T., Canavan, J., & Forkan, C. (2010). *Big Brothers Big Sisters of Ireland: Evaluation study. Report 2: Qualitative evidence*. Child and Family Research Centre, National University of Ireland.
- Dolan, P., Brady, B., O'Regan, C., Russell, D., Canavan, J., & Forkan, C. (2010). *Big Brothers Big Sisters (BBBS) of Ireland: Evaluation study. Report 1: Randomised control trial and implementation report*. Galway: Child and Family Research Centre, National University of Ireland.
- DuBois, D. L., & Friend, D. (2017). Big Brothers Big Sisters. In K. Pepper (Ed.), *The SAGE encyclopedia of out-of-school learning* (pp. 63–65). SAGE Publications Inc.
- 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., & Keller, T. E. (2017). Investigation of the integration of supports for youth thriving into a community-based mentoring program. *Child Development*, 88(5), 1480–1491. <https://doi.org/10.1111/cdev.12887>
- 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., & 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>
- Eddy, J. M., Martinez, C. R., Grossman, J. B., Cearley, J. J., Herrera, D., Wheeler, A. C., ... Seely, J. R. (2017). A randomized controlled trial of a long-term professional mentoring program for children at risk: Outcomes across the first 5 years. *Prevention Science*, 18(8), 899–910. <https://doi.org/10.1007/s11121-017-0795-z>
- Ferro, A., Wells, S., Speechley, K. N., Lipman, E., & DeWit, D. (2014). The measurement properties of mentoring relationship quality scales for mentoring programs. *Prevention Science*, 15(5), 663–673. <https://doi.org/10.1007/s11121-013-0392-8>
- Festinger, L. (1957). *A theory of cognitive dissonance* (Vol. 2). Stanford University Press.
- 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. <https://doi.org/10.13140/RG.2.2.18166.70728>
- Goense, P. B., Assink, M., Stams, G. J., Boendermaker, L., & Hoeve, M. (2016). Making “what works” work: A meta-analytic study of the effect of treatment integrity on outcomes of evidence-based interventions for juveniles with antisocial behavior. *Aggression and Violent Behavior*, 31, 106–115. <https://doi.org/10.1016/j.avb.2016.08.003>
- Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38(5), 581–586. <https://doi.org/10.1111/j.1469-7610.1997.tb01545.x>
- Groenman, A. P., Janssen, T. W., & Oosterlaan, J. (2017). Childhood psychiatric disorders as risk factor for subsequent substance abuse: A meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, 56(7), 556–569. <https://doi.org/10.1016/j.jaac.2017.05.004>
- Grossman, J. B., & Johnson, A. (1999). Assessing the effectiveness of mentoring programs. In J. B. Grossman (Ed.), *Contemporary issues in mentoring* (pp. 24–47). Public/Private Ventures. <https://ppv.issuelab.org/resource/contemporary-issues-in-mentoring.html>
- 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/10.1023/a:1014680827552>
- Guryan, J., Christenson, S., Cureton, A., Lai, L., Ludwig, J., Schwarz, C., ... Turner, M. C. (2021). The effect of mentoring on school attendance and academic outcomes: A randomized evaluation of the Check & Connect Program. *Journal of Policy Analysis and Management*, 40(3), 841–882. <https://doi.org/10.2139/ssrn.3670496>
- Harter, S. 1985. *Manual for the self-perception profile for children*. University of Denver.
- Heppe, E. C., Willemsen, A. M., Kef, S., & Schuengel, C. (2021). Evaluation of a community-based mentoring program on psychosocial functioning of adolescents with a visual impairment: A randomized controlled trial. *British Journal of Visual Impairment*, 39(3), 251–265. <https://doi.org/10.1177/0264619620935944>
- Herrera, C., DuBois, D. L., & Grossman, J. B. (2013). The role of risk: Mentoring experiences and outcomes for youth with varying risk profiles. *A Public/Private Ventures project distributed by MDRC*. <https://www.mdrc.org/sites/default/files/Role%20of%20Risk%20Final-web%20PDF.pdf>
- Herrera, C., Sipe, C. L., & McClanahan, W. S. with Arbreton, A. J. A., & Pepper, S. K. (2000). *Mentoring school-age children: Relationship development in community-based and school-based programs*. Public/Private Ventures. <https://ppv.issuelab.org/resource/mentoring-school-age-children-relationship-development-in-community-based-and-school-based-programs.html>
- Herrera, C., Grossman, J. B., Kauh, T. J., Feldman, A. F., & McMaken, J. with Jucovy, L. Z. (2007). *Making a difference in schools: The Big Brothers Big Sisters school-based mentoring impact study*. Public/Private Ventures. <http://ppv.issuelab.org/resources/574/574.pdf>
- ICF International (2011). *Mentoring children affected by incarceration: An evaluation of the Amachi Texas Program*. Office of Juvenile Justice and Delinquency Prevention.



- <https://www.evidencebasedmentoring.org/wp-content/uploads/2013/11/RCT-Study-of-Amachi-Texas-Evaluation-Report-December-2011.pdf>.
- Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials – A practical guide with flowcharts. *BMC Medical Research Methodology*, 17(2), Article 162. doi: 10.1186/s12874-017-0442-1.
- Jarjoura, G. R., Tanyu, M., Forbush, J., Herrera, C., & Keller, T. E. (2018). *Evaluation of the mentoring enhancement demonstration program: Technical report*. Office of Justice Programs' National Criminal Justice Reference Service Document Number 252167. <https://www.ncjrs.gov/pdffiles1/ojdp/grants/252167.pdf>.
- Ji, P., Flay, B. R., & DuBois, D. L. (2013). Social-emotional and character development scale: Development and initial validation with urban elementary school students. *Journal of Character Education*, 9(2), 121.
- Kanchewa, S. S., Yoviene, L. A., Schwartz, S. E., Herrera, C., & Rhodes, J. E. (2018). Relational experiences in school-based mentoring: The mediating role of rejection sensitivity. *Youth & Society*, 50(8), 1078–1099. <https://doi.org/10.1177/0044118X16653534>
- Karcher, M. J. (2005). The effects of developmental mentoring and high school mentors' attendance on their younger mentees' self-esteem, social skills, and connectedness. *Psychology in the Schools*, 42(1), 65–77. <https://doi.org/10.1002/pits.20025>
- Keller, T. E., Drew, A. L., Clark-Shim, H., Spencer, R., & Herrera, C. (2020). It's about time: Staff support contacts and mentor volunteer experiences. *Journal of Youth Development*, 15(4), 145–161. <https://doi.org/10.5195/jyd.2020.879>
- Kerr, D. C. R., & King, C. A. (2014). Youth with mental health needs. In D. L. DuBois, & M. J. Karcher (Eds.), *Handbook of youth mentoring* (2nd ed., pp. 325–340). SAGE Publications Inc.
- King, C. A., Gipson, P. Y., Arango, A., Lernihan, D., Clark, M., Ewell Foster, C., ... Stone, D. (2021). LET's CONNECT community mentorship program for adolescents with peer social problems: A randomized intervention trial. *American Journal of Community Psychology*, 68(3–4), 310–322. <https://doi.org/10.1002/ajcp.12528>
- Kuperminc, G. P., & Thomason, J. D. (2014). Group mentoring. In D. L. DuBois, & M. J. Karcher (Eds.), *Handbook of youth mentoring* (2nd ed., pp. 273–290). SAGE Publications Inc.
- Larose, S., & Tarabulsy, G. M. (2005). Academically at-risk students. In D. L. DuBois, & M. J. Karcher (Eds.), *Handbook of youth mentoring* (pp. 440–453). SAGE Publications Inc.
- Lewis, K. M., Vuchinich, S., Ji, P., DuBois, D. L., Acock, A., Bavarian, N., ... Flay, B. R. (2016). Effects of the Positive Action program on indicators of positive youth development among urban youth. *Applied Developmental Science*, 20(1), 16–28. <https://doi.org/10.1080/10888691.2015.1039123>
- Lindstrom Johnson, S., Jones, V., Ryan, L., DuBois, D. L., Fein, J. A., & Cheng, T. L. (2022). Investigating effects of mentoring for youth with assault injuries: Results of a randomized-controlled trial. *Prevention Science. Advance online publication*. <https://doi.org/10.1007/s11121-022-01406-z>
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE Publications, Inc. <https://psycnet.apa.org/record/2000-16602-000>.
- Liu-Seifert, H., Andersen, S., Case, M., Sparks, J., Holdridge, K. C., Wessels, A. M., ... Siemers, E. (2017). Statistical properties of continuous composite scales and implications for drug development. *Journal of Biopharmaceutical Statistics*, 27(6), 1104–1114. <https://doi.org/10.1080/10543406.2017.1315819>
- Malhotra, S., & Das, P. P. (2007). Understanding childhood depression. *Indian Journal of Medical Research*, 125, 115–128.
- Millenky, M., Schwartz, S. E., & Rhodes, J. E. (2014). Supporting the transition to adulthood among high school dropouts: An impact study of the National Guard Youth ChalleNGe Program. *Prevention Science*, 15(4), 448–459. <https://doi.org/10.1007/s11121-013-0388-4>
- Peaslee, L., & Teye, A. C. (2015). *Testing the impact of mentor training and peer support on the quality of mentor-mentee relationships and outcomes for at-risk youth*. Department of Justice: U.S. <https://ojdp.ojp.gov/library/publications/testing-impact-mentor-training-and-peer-support-quality-mentor-mentee>.
- Porzig, R. (June, 2021). *2020 Big Brothers Big Sisters of America annual impact report*. Big Brothers Big Sisters of America. <https://www.bbbs.org/wp-content/uploads/2020-Big-Brothers-Big-Sisters-of-America-Annual-Impact-Report.pdf>.
- Posner, J. K., & Vandell, D. L. (1994). Low-income children's after-school care: Are there beneficial effects of after-school programs? *Child Development*, 65(2), 440–456. <https://doi.org/10.2307/1131395>
- Prinstein, M. J., & La Greca, A. M. (2009). Childhood depressive symptoms and adolescent cigarette use: A six-year longitudinal study controlling for peer relations correlates. *Health Psychology*, 28(3), 283–291. <https://doi.org/10.1037/a0013949>
- Raposa, E. B., Rhodes, J., Stams, G. J. J. M., Card, N., Burton, S., Schwartz, S., ... Hussain, S. (2019). The effects of youth mentoring programs: A meta-analysis of outcome studies. *Journal of Youth and Adolescence*, 48(3), 423–443. <https://doi.org/10.1007/s10964-019-00982-8>
- Resnjanskij, S., Ruhose, J., Wiederhold, S., & Woessmann, L. (2021). *Can mentoring alleviate family disadvantage in adolescence? A field experiment to improve labor-market prospects*, Discussion Paper, No. 277, Ludwig-Maximilians-Universität München und Humboldt-Universität zu Berlin, Collaborative Research Center Transregio 190 - Rationality and Competition, München und Berlin. <http://hdl.handle.net/10419/233500>.
- Rhodes, J. E. (2005). A model of youth mentoring. In D. L. DuBois, & M. J. Karcher (Eds.), *Handbook of youth mentoring* (pp. 30–43). SAGE Publications Inc.
- Rhodes, J. E., Grossman, J. B., & Resch, N. R. (2000). Agents of change: Pathways through which mentoring relationships influence adolescents' academic adjustment. *Child Development*, 71(6), 1662–1671. <https://doi.org/10.1111/1467-8624.00256>
- Rhodes, J. E., Reddy, R., Roffman, J., & Grossman, J. B. (2005). Promoting successful youth mentoring relationships: A preliminary screening questionnaire. *Journal of Primary Prevention*, 26(2), 147–167. <https://doi.org/10.1007/s10935-005-1849-8>
- Sánchez-Meca, J., Marín-Martínez, F., & Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods*, 8(4), 448–467. <https://doi.org/10.1037/1082-989X.8.4.448>
- Saraceno, L., Heron, J., Munafò, M., Craddock, N., & van den Bree, M. B. (2012). The relationship between childhood depressive symptoms and problem alcohol use in early adolescence: Findings from a large longitudinal population-based study. *Addiction*, 107(3), 567–577. <https://doi.org/10.1111/j.1360-0443.2011.03662.x>
- Stone, L. L., Otten, R., Engels, R. C., Vermulst, A. A., & Janssens, J. M. (2010). Psychometric properties of the parent and teacher versions of the Strengths and Difficulties Questionnaire for 4- to 12-year-olds: A review. *Clinical Child and Family Psychology Review*, 13(3), 254–274. <https://doi.org/10.1007/s10567-010-0071-2>
- Sullivan, T. R., White, I. R., Salter, A. B., Ryan, P., & Lee, K. J. (2018). Should multiple imputation be the method of choice for handling missing data in randomized trials? *Statistical Methods in Medical Research*, 27(9), 2610–2626. <https://doi.org/10.1177/0962280216683570>
- Tierney, J. P., Grossman, J. B., with Resch, N. L. (1995) *Making a difference: An impact study of Big Brothers Big Sisters*. Public/Private Ventures. <http://ppv.issueab.org/resources/11972/11972.pdf>.
- Tolan, P. H., Henry, D. B., Schoeny, M. S., Lovegrove, P., & Nichols, E. (2014). Mentoring programs to affect delinquency and associated outcomes of youth at risk: A comprehensive meta-analytic review. *Journal of Experimental Criminology*, 10(2), 179–206. <https://doi.org/10.4073/csr.2013.10>
- Turner, N., Joinson, C., Peters, T. J., Wiles, N., & Lewis, G. (2014). Validity of the Short Mood and Feelings Questionnaire in late adolescence. *Psychological Assessment*, 26(3), 752–762. <https://doi.org/10.1037/a0036572>
- Wheeler, M. E., & DuBois, D. L. (2009). *Analyses of responses to agency practices survey for Big Brothers Big Sisters of America's community-based mentoring program*. Unpublished report.
- World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. World Health Organization. <https://apps.who.int/iris/handle/10665/37958>.
- Yeager, D. S., & Krosnick, J. A. (2011). Does mentioning "some people" and "other people" in a survey question increase the accuracy of adolescents' self-reports? *Developmental Psychology*, 47(6), 1674–1679. <https://doi.org/10.1037/a0025440>