School-Based Mentoring Programs: Using Volunteers to Improve the Academic Outcomes of Underserved Students

Bayer, Amanda and Grossman, Jean and DuBois, David

August 2013

School-Based Mentoring Programs:

Using Volunteers to Improve the Academic Outcomes of Underserved Students

Amanda Bayer
(Swarthmore College)

Jean Baldwin Grossman
(Princeton University and MDRC)

and

David L. DuBois
(University of Illinois at Chicago)

August 2013

Correspondence to:

Amanda Bayer
Department of Economics
Swarthmore College
500 College Avenue
Swarthmore PA 19081-1397
610-328-7821
abayer1@swarthmore.edu

Acknowledgments

We are grateful to the organizations that funded and supported the original evaluation providing the data for the present study, including The Atlantic Philanthropies, Philip Morris USA, The William T. Grant Foundation, The Edna McConnell Clark Foundation, and Big Brothers Big Sisters of America. We also thank Lynn Conell-Price for research assistance, Joshua Malbin for editing, and Carla Herrera for her support and wisdom.

Dissemination of MDRC publications is supported by the following funders that help finance MDRC's public policy outreach and expanding efforts to communicate the results and implications of our work to policymakers, practitioners, and others: The Annie E. Casey Foundation, The George Gund Foundation, Sandler Foundation, and The Starr Foundation.

In addition, earnings from the MDRC Endowment help sustain our dissemination efforts. Contributors to the MDRC Endowment include Alcoa Foundation, The Ambrose Monell Foundation, Anheuser-Busch Foundation, Bristol-Myers Squibb Foundation, Charles Stewart Mott Foundation, Ford Foundation, The George Gund Foundation, The Grable Foundation, The Lizabeth and Frank Newman Charitable Foundation, The New York Times Company Foundation, Jan Nicholson, Paul H. O'Neill Charitable Foundation, John S. Reed, Sandler Foundation, and The Stupski Family Fund, as well as other individual contributors.

The findings and conclusions in this report do not necessarily represent the official positions or policies of the funders.

Abstract

Previous research suggests that school-based mentoring programs like those offered by Big Brothers Big Sisters of America (BBBSA) yield small but statistically significant improvements in the academic performance of mentored students and in their beliefs in their own scholastic efficacy. The present study uses data from a randomized control trial involving over 1,000 students from 71 schools across the country to investigate further the academic benefits of school-based mentoring, and to enrich the field's understanding of how schools can use volunteers to support students. We employ instrumental variables and other approaches to provide insight into why the BBBSA school-based mentoring program is effective, finding that the relationship between mentor and protégé appears to play a key role. The evidence suggests that developing a close relationship with a mentor led to better academic outcomes for students; in contrast, students who were mentored but did not experience a close relationship showed no improvement in academic outcomes relative to the control group. This pattern holds for mentoring relationships of various durations. In addition, there is no evidence that mentoring programs with an academic focus produced better academic outcomes than relationship-only programs. Findings do reveal, however, that programs structured with weekly meetings and with opportunities for pairs to interact outside of a large-group setting were more likely to generate close mentor-protégé relationships. Beyond reporting new empirical findings, this paper contributes a theoretical structure with which to assess the results of randomized evaluations of mentoring programs.

Contents

List of Exhibits

Table

Figure

I. Introduction

Financially distressed school districts need to identify low-cost ways to support struggling students in struggling schools. School-based mentoring programs (SBM), in which volunteers meet regularly with students on school grounds, are an increasingly popular option. As of 2005, the last date for which data are available, SBM was the fastest-growing form of mentoring in the United States, with 870,000 adults mentoring children in schools as part of one formal program or another (MENTOR, 2006).

The mission of mentoring programs such as those offered by Big Brothers Big Sisters of America (BBBSA) is to "provide children facing adversity with ... one-to-one relationships that change their lives for the better" (Big Brothers Big Sisters of America, 2013c). Even when based in schools, the primary objective of the mentoring program is to provide a friend, not a tutor, as illustrated by the following description from the organization's website:

> Although it takes place at schools, our School-Based Mentoring program isn't limited to the classroom. Of course, some Littles do talk with their Bigs about class, or do homework, or read together, but it's perfectly fine to shoot hoops in the gym or play on the playground. At the end of the day, it's really all about starting a friendship, providing guidance and inspiring them to reach their potential. (Big Brothers Big Sisters of America, 2013b)

The overarching aim of this paper is to enrich the field's understanding of how volunteer mentors can best support the academic mission of schools. Our central empirical analysis investigates whether emotionally closer relationships between mentors and protégés lead to better academic outcomes. We wish to determine whether SBM works primarily through the connection that the protégé feels with the mentor rather than through other, more direct processes, such as engaging the child in academic activities during meetings.

Prior research on mentoring relationships outside of school does point toward relationship closeness (for example, DuBois & Neville, 1997; Parra et al., 2002) and related indicators of the emotional quality of the mentor-protégé tie (for example, Thomson & Zand, 2010; Zand et al., 2009) as important influences on youth outcomes. There is preliminary evidence that this may also be the case for SBM, or at least that closeness promotes protégé and mentor perceptions of relationship quality (Pryce & Keller, 2012). To date, however, there has not been a rigorous test of the role of relationship closeness in promoting academic outcomes in SBM. The present research addresses this gap using data from a randomized trial evaluation of the BBBSA SBM program conducted by Public/Private Ventures (P/PV). Because of course relationship closeness cannot be randomly assigned, we use instrumental variables and other controls to

account for factors such as the preexisting characteristics of protégés that may contribute to both relationship closeness and youth outcomes.

Prior work with the same data set examined the effects on academic outcomes of match length and status (that is, whether the original match was intact and, if not, whether the protégé had been rematched with a new mentor). That work found negative impacts for protégés given a new mentor after a prior match was terminated (Grossman, Chan, Schwartz, & Rhodes, 2012). Theoretically, though, a positive emotional quality or tenor in a mentoring relationship should mediate the benefits of program participation, independent of whatever effects may be associated with relationship length or status (Rhodes, 2005). Accordingly, in our analyses we explore whether relationship quality matters even when relationship length and current status are taken into account.

Finally, we investigate the characteristics of programs, mentors, and protégés that may increase the likelihood of a close relationship. We present findings that may help practitioners design SBM programs more conducive to close mentor-protégé relationships, and perhaps to more positive academic outcomes.

The next section provides background information on youth mentoring practice and research. The third section of the paper summarizes the experimental design and data, followed by a discussion of our analytical method. Section IV presents our results, and Section V offers a summary and discussion.

II. School-Based Mentoring: Background

Mentoring is a youth development intervention that seeks to provide a child with an additional caring adult relationship. Mentoring programs pair children with volunteer adults or older students and facilitate regular meetings of the pairs. The theory is that the child benefits from the additional adult support and guidance. Mentoring organizations like Big Brothers Big Sisters target children in risky circumstances, including "children living in single parent homes, growing up in poverty, and coping with parental incarceration" (Big Brothers Big Sisters of America, 2013a).

In the school-based form of mentoring, adults or older students meet with their protégés on school grounds during the school day or immediately after, typically for one hour per week during the academic year. BBBSA began partnering with schools over fifteen years ago to extend the reach of its successful community-based mentoring (CBM) services. Although BBBSA screens, trains, and supervises mentors in both types of programs, the contact and context of SBM is more constrained than in the community-based version. CBM, as practiced by BBBSA and other organizations, typically involves longer meetings, and locations and activities that the pairs choose themselves. The more limited time commitment and firmer structure of SBM make it easier for agencies to recruit volunteers. And because protégé referrals to SBM programs come from teachers and school personnel rather than being dependent on parental initiative as in CBM, SBM also reaches groups of children who would not otherwise be served (Herrera, 2004). Although there are other providers of SBM, BBBSA is the largest in the United States. It alone administered 126,000 school-based matches in 2006 (Herrera et al., 2007).

Available evidence on the effectiveness of SBM from randomized control trials is encouraging, though somewhat mixed. The P/PV random assignment impact evaluation of the BBBSA SBM program that provides the data for the present research found that students who were assigned to receive mentoring showed small but statistically significant improvements in several academic outcomes, including teacher-rated academic performance and student-reported feelings of scholastic efficacy (Herrera et al., 2007). On the other hand, an evaluation of SBM programs funded by the U.S. Department of Education's Student Mentoring Program concluded that the program did not have significant effects on academic outcomes (Bernstein et al., 2009). In considering the results of this latter evaluation, Wheeler, Keller, and DuBois (2010) noted that its findings "may have been attenuated by a range of factors such as the absence of a well-delineated program model, a lack of additional implementation support for study agencies, and a relatively high percentage of the children in the intervention group who were never paired with a mentor."

In contrast, the BBBSA SBM program has well-defined national standards. These include, for example, an expectation that mentors receive monthly support contacts from program staff, a factor that research indicates is important for establishing high-quality mentoring relationships (Herrera, DuBois, & Grossman, 2013). Of further note is that the agencies participating in the P/PV evaluation of the BBBSA SBM program were required to have at least four years of experience delivering school-based mentoring, strong agency leadership, and strong, established relationships with participating schools (Herrera et al., 2007). Moreover, in this evaluation only 7 percent of the children in the intervention group were never matched with a mentor, less than half the rate in the U.S. Department of Education study. For these reasons, analysis of the data from the P/PV evaluation offers an especially rigorous and informative basis for examining mechanisms underlying the effects of SBM. Furthermore, agencies involved in this trial had the opportunity to vary potentially consequential aspects of program practice, including the extent to which academics were a focus during meetings between mentors and protégés. We exploit this variation to explore whether some program practices are more or less conducive to close relationships between protégés and mentors.

Theoretical models developed in the context of CBM programs suggest that the critical component of mentoring is the relationship that develops between the child and the mentor (Rhodes, 2005). Such models typically assume that if a supportive adult consistently spends time with a child, the pair will form an emotionally close and mutually trusting relationship. Thanks to this relationship the child can begin to develop more positive relationships with teachers, parents, and peers, and begin to think more positively about him- or herself (Harter, 1990), ultimately leading to better outcomes in areas such as academics. Available research backs up these theoretical models: relationship quality, including feelings of closeness specifically, does have an important influence on youth outcomes.

Most of this research, however, has been on mentoring outside of school. There are several reasons it may not automatically be possible to generalize its conclusions to SBM. For example, mentoring relationships are typically shorter in SBM programs than in CBM programs. Those relationships might not have a chance to reach the levels of closeness or emotional connection needed to promote better academic outcomes (by, for example, radiating effects on the protégé's relationships with teachers and parents).

On somewhat different grounds, Cavell and Elledge (forthcoming) also challenge the assumption that a good relationship is a precondition for positive youth outcomes, which they refer to as the "mentoring-as-relationship" perspective. They argue that mentoring also may be influential because it provides a context for activities or experiences that serve specific goals, such as the development of improved academic skills. From this alternative "mentoring-as-context" perspective, they note that benefits to protégés would be expected to accrue independent of their mentoring relationships' strength and duration. Because SBM programs are by

definition based in schools, and because some SBM programs specifically emphasize academic activities, mentoring-as-context mechanisms could end up accounting for the academic benefits of program participation. The present investigation is thus well suited to inform the debate on this important, unresolved issue.

In their investigation using data from the same P/PV study, Grossman et al. (2012) find little evidence that program impacts on academic outcomes varied according to the length of time a protégé was matched with a mentor, but they do find variation among protégés based on match status. Relative to controls, protégés in intact mentoring relationships showed significant improvement on teacher-rated academic performance. On the other hand, those who had been rematched with a new mentor after an initial match ended did significantly worse than the controls. These negative effects were not observed for protégés whose initial mentoring relationships had ended but who had not been rematched. Similarly, despite the findings of Grossman et al., prior research with CBM programs has supported the hypothesis that mentoring relationships and their effects grow stronger over time. As we investigate whether relationship closeness is the key ingredient to an effective school-based mentoring experience, we therefore take care to separate the effects of closeness from the potentially confounding effects of both match length and match status.

It is also useful to consider briefly the evidence on a broader array of school-based prevention programs, as they too introduce caring adults into students' lives. A meta-analysis of evaluations of social and emotional learning programs in schools (Durlak et al., 2011) found that program effects on academic and other outcomes were substantially greater when programs were structured to engage students in systematic skill-building activities (that is, "SAFE": Sequenced, Active, Focused, and Explicit). On the surface, this finding would seem to suggest that a close relationship with a mentor is not necessarily the key to better academic outcomes. Yet SBM programs specifically focus on cultivating emotional connections between children and adults, so they may capitalize on relational factors to foster academic improvements among participants.

III. Method

Study Design and Sample

The sample for our study consists of the students who participated in the randomized control trial of the Big Brothers Big Sisters of America (BBBSA) SBM program during the 2004-2005 school year (Herrera et al., 2007). Study participants were recruited by 10 BBBSA study agencies across the country, each with four or more years of experience in SBM. Recruitment proceeded in the usual way, that is, mostly through referrals by school staff; children had to be entering grades four through nine and had to have parental consent to be part of the research. Over a thousand (1,139) students across 71 public schools in rural and urban school districts (41 elementary, 27 middle, and 3 high schools) met these criteria. After completing a baseline survey, half (565) of these students were randomly assigned by researchers to the treatment group to be matched with a mentor and the other half (574) were placed on a waiting list to be matched 15 months later, at the end of the study.[1] Randomization was stratified by school. The students' teachers and mentors also completed baseline surveys in the fall of 2004. Follow-up surveying of all groups occurred at the end of the school year, beginning in April 2005.

Table 1 reports basic descriptive statistics for the control and treatment samples at baseline. Overall, just over half (54 percent) of the students were female and the majority (63 percent) were members of ethnic and racial minority groups. Most (69 percent) received free or reduced-price lunches. On average they struggled academically, receiving below a 3.0 on the teacher survey (the rating for "average" or "satisfactory" performance). The average age was 11. As shown in Table 1, there were no statistically significant differences between the control group and the group assigned to treatment, indicating successful randomization.

All but 6 percent of the protégés completed the survey at follow up. Survey completion rates were slightly lower for teachers, with follow-up outcome measures missing for approximately 16 percent of the overall sample.[2] Importantly, however, survey noncompletion rates are almost identical for the treatment and control groups (6 percent versus 7 percent for the student survey, and 16 percent versus 17 percent for the teacher survey). According to evidence standards developed and used by the What Works Clearinghouse of the U.S. Department of Education's Institute of Education Sciences (2011), this combination of overall and differential attrition should result in an acceptable level of bias even under conservative assumptions.

[1]There were originally 575 control students but one dropped out of the study. Students who were referred by Child Protective Services were exempt from the study lottery but were not included in the study.

[2]Analyses indicate that students with lower baseline academic performance were slightly less likely to complete the survey, so the baseline academic measure is included as a covariate in all specifications.

Table 1

Student Characteristics at Baseline

	Control		Treatment		Difference (Control-Treatment)	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard error
Age	11.22	1.66	11.24	1.67	-.02	.099
African-American/Hispanic (proportion)	0.61	0.49	0.64	0.48	-.04	.029
Female (proportion)	0.54	0.50	0.54	0.50	.00	.030
Free/reduced-price lunch (proportion)	0.69	0.43	0.69	0.43	-.00	.025
Teacher-rated outcome measures:						
Overall academic performance	2.47	1.09	2.56	1.10	-.08	.070
Quality of schoolwork	2.82	0.91	2.84	0.94	-.02	.058
Completion of schoolwork	2.96	0.99	3.00	1.06	-.04	.065
Student-reported outcome measure:						
Scholastic efficacy	2.75	0.64	2.80	0.62	-.05	.038
Sample size	574		565			

NOTES: See Section III, "Outcome Measures and Key Variables," for an explanation of how outcome measures were derived. A two-tailed t-test was applied to differences between the groups. No differences were statistically significant at a 10 percent level.

Because SBM does not require same-gender matches (most CBM programs do), the percentage of female mentors was greater than the percentage of female students. In fact, mentors were predominately female (72 percent). More than three-quarters (77 percent) of the mentors were white, and nearly half (48 percent) were high school students.

Program Overview

About half (49 percent) of SBM programs had protégés meet their mentors only during the school day, with the others meeting after school (47 percent) or both during and after school (4 percent). Programs reported using a variety of locations in the school for meetings, including the cafeteria (41 percent), library (34 percent), and a designated classroom (33 percent). Some programs used more than one meeting place. Almost half of all mentoring pairs shared their meeting space with other pairs.

Table 2 presents some statistics describing the features of the programs and characteristics of the mentors in this study. These variables are used in the analyses presented in Section IV, "Facilitating Relationships in School-Based Mentoring," which explore the characteristics of programs and mentors that increase the likelihood of a close relationship. Mentors reported engaging in a wide variety of activities with their protégés, most of which were chosen by

protégé and mentor together (49 percent) or by the protégé alone (20 percent). For example, mentors reported doing each of the following types of activities "a lot": talking casually (71 percent), talking about family or friends (43 percent and 44 percent, respectively), talking about the future (30 percent), playing indoor games (54 percent), doing creative activities (36 percent), playing sports (25 percent), doing homework (27 percent), and talking about academic issues (31 percent). Many pairs met for between 45 and 60 minutes (40 percent) but another 39 percent met for more than an hour. Almost all matches met three (10 percent) or four (79 percent) times a month. Over the year, the average protégé received 17 hours of mentoring over 5.3 months.

On average, the BBBSA agencies overseeing the school programs had 9.5 years of experience implementing SBM. BBBSA supported sites during implementation with meetings and teleconferences. Prematch mentor training lasted about 45 minutes on average.

Table 2

Program Features

Treatment	Mean	Standard Deviation
Academically focused program[a]	0.40	0.49
All pairs meeting at same time and location[a]	0.44	0.50
Some structured time[a]	0.71	0.46
Infrequent meetings[a]	0.13	0.33
Volunteer recognition[b]	2.63	1.49
High school student mentor[a]	0.49	0.50
College student mentor[a]	0.16	0.37
Same gender mentor as student[a]	0.80	0.39
Same race mentor as student[a]	0.49	0.50
"Mentor shows up when he/she is supposed to"[c]	3.33	0.94
Amount of individual mentor training[d]	3.10	1.45
Sample size	565	

NOTES: [a]The mean of this variable represents the proportion of students assigned to treatment who were mentored under the specified condition.

[b]This variable records the number of ways programs recognized volunteers' efforts.

[c]The mean of this variable represents the average rating by the student on a scale from 1 = "not at all true" to 5 = "very true."

[d]This variable is the sum of mentor training before and after match using the following categorical scale: 1 = less than 30 minutes, 2 = 30 minutes to less than one hour, 3 = one hour to less than two hours, 4 = two hours to less than 4 hours, and 5 = four hours or more.

Outcome Measures and Key Variables

This paper examines changes in four academic measures, each measured at baseline and at follow-up for each student:

- **Overall academic performance** is a rating by the student's teacher of the student's "overall academic performance" on a single-item, five-point scale from 1 = "below grade level" to 5 = "excellent" (Pierce, Hamm, & Vandell, 1999).
- **Quality of work** is a three-item scale score calculated as a mean of teacher ratings of the correctness, neatness, and completeness of a child's classwork on a five-point scale from 1 = "well below average" to 5 = "well above average" (Herrera, 2004; alpha values of 0.88 at baseline and 0.90 at follow-up indicate internal reliability (Cronbach, 1951)).
- **Completion of schoolwork** is a two-item average of the teacher's report of how often the student completes in-class and homework assignments on a five-point scale ranging from 1 = "well below average" to 5 = "well above average" (Herrera, 2004; alpha values of 0.93 at baseline and 0.94 at follow-up indicate internal reliability).
- **Scholastic efficacy** is the average across six items of a student's ratings of his or her own ability to do schoolwork on a four-point scale from 1 = "not at all true" to 4 = "very true." Items include "I do very well at my class work" and a reverse coding of "I often forget what I learn" (Harter, 1985; alpha values of 0.70 at baseline and 0.72 at follow-up indicate internal reliability).

In all cases, a higher score indicates better outcomes.

Four explanatory variables are central to our exploration of the mechanism of SBM:

- **Closeness** is a single-item measure in which protégés were asked to respond to the question "How close do you feel to your mentor?" on a four-point scale where 1 = "not close at all," 2 = "not very close," 3 = "somewhat close," and 4 = "very close." This question was part of the follow-up survey completed by protégés at the end of the school year, with students instructed to rate their current or most recent mentor.
- **Match status** is the status of the protégé's original mentoring match at the end of the school year. The original mentoring relationship was "intact," or that relationship ended and the student experienced a "rematch" with a new

mentor, or the relationship ended and the student was given "no rematch" with a new mentor.

- **Match length** is the total number of weeks of a protégé's current or most recent mentoring match (corresponding to the relationship for which we have closeness ratings) at the time of the follow-up survey.[3]

- **Teacher relationship** is the average across 11 items of a student's ratings of teacher connectedness (Karcher, 2003) and teacher relationship quality. Items are rated on a four-point scale from 1 = "not at all true" to 4 = "very true," and include "I always try hard to earn my teachers' trust" and a reverse coding of "My teachers don't know me very well this year" (alpha values of 0.82 at baseline and 0.82 at follow-up indicate internal reliability). We use the baseline value of this variable as a measure of the child's ability to have a close relationship with a nonparental adult.

Analysis Plan

Regression specifications in which the outcome, Y_i, is regressed on the assigned treatment indicator, Z_i, provide unbiased estimates of the average impact of SBM across all protégés. Consider:

$$Y_i = \alpha_1 + \beta_{ITT} Z_i + X_i \gamma_1 + \varepsilon_i \qquad (1)$$

where X_i is a vector of covariates, which in our analyses is the pretreatment value of the outcome, and ε_i is the unexplained variation in outcomes. These intention-to-treat (ITT) estimates, β_{ITT}, help inform us about the average effectiveness of the programs as currently configured.

β_{ITT} provides a fairly good estimate of the impact of mentoring-as-context, since only 7 percent of the treatment students were never matched. But because only 75 percent of the students assigned to treatment rated their relationship with their mentor as "somewhat close" or "very close," it does not represent the impact of providing a child with a close, caring relationship with a nonparental adult, widely viewed as the essence of mentoring. The "partial compliance" of treatment students to forming a close relationship means that β_{ITT} does not equal the effect of the "close adult relationship" treatment. The randomized assignment, Z_i, only influ-

[3]Grossman et al. (2012) use a definition of match length that sums all days of mentoring across all matches. We define match length slightly differently, given our focus on the relationship between the two individuals in the match.

ences the probability that a student in the mentoring group receives the treatment of interest, T_i, where T_i represents a close mentor-protégé relationship.[4]

We therefore estimate the causal effect of having a close relationship with a mentor, that is, the effect among the "compliers" with the SBM treatments, à la Angrist, Imbens, and Rubin (1996). Regressions in which outcomes are regressed on the receipt of the intended treatment, T_i, provide the treatment-on-treated (TOT) estimates, β_{TOT}.

$$Y_i = \alpha_2 + \beta_{TOT} T_i + X_i \gamma_2 + \varepsilon_i \qquad (2)$$

With an unbiased estimate of β_{TOT}, we can gauge the impact of SBM when it is operating as designed, providing children with close relationships with program mentors.

In producing our estimates, we must take care not to reintroduce selection bias into the results. Although treatment status, Z_i, was assigned randomly, the students who actually received the intended treatment, T_i, are not necessarily a random subset. These students may have an underlying characteristic, A_i, affecting both their ability to form a relationship with a mentor and their academic progress over the study year — for example, a general ability to have good relationships with adults. One approach is to control for the characteristic explicitly. The youth survey includes questions about a student's relationship with his or her teacher, and we use baseline responses to construct a variable A_i, which provides a direct measure of a student's ability to have a close relationship with a nonparental adult, to include in our ordinary least squares (OLS) regressions.

$$Y_i = \alpha_3 + \beta_3 T_i + \rho A_i + X_i \gamma_3 + \varepsilon_i \qquad (3)$$

Alternatively, when compliance is imperfect, we can use the randomized treatment status, Z_i, as an instrument for the treatment of interest, having a relationship with a mentor, T_i. As established in Angrist, Imbens, and Rubin (1996), random assignment status is an ideal candidate for an instrument because it is independent of student characteristics. We thus derive the Wald estimate of the causal effect among compliers: the difference in outcomes between the nominal treatment and control groups divided by the difference in the probability of receiving treatment between the groups. This estimate, which provides the effect of the treatment on the treated or, more generally, the local average treatment effect, measures the effect of program participation on those protégés who successfully connected with their assigned mentors (Bloom, 1984; Imbens & Angrist, 1994). Specifically,

[4] Although we know that $E[Y_i(0)|Z = 1] - E[Y_i(0)|Z = 0]$ is equal to zero due to random assignment, the intention to treat estimate, $E[Y_i|Z = 1] - E[Y_i|Z = 0]$, is equal to the causal effect of Z, not the effect of the treatment, T.

$$(E[Y_i|Z = 1] - E[Y_i|Z = 0]) / Pr[T_i = 1|Z = 1] \quad (4)$$

Two-stage least squares (2SLS) is used to estimate this effect of the treatment on the treated, where Z_i acts as an instrument for T_i.

This instrumental variables estimate can be interpreted as the average treatment effect for the compliers — that is, the effect on the students who form a close relationship with a program mentor — under two assumptions: monotonicity and independence (Angrist, Imbens, & Rubin, 1996). Monotonicity requires that the assignment of a mentor increase the likelihood of a new close mentoring relationship for each and every student assigned to treatment. This requirement is met, since control students were not offered the opportunity to form relationships with BBBSA mentors.

Independence requires two things: first, that we be able to identify the causal impact of the instrument, which is also true by construction here, and second, that potential outcomes not be directly affected by the instrument. This last requirement, also known as an exclusion restriction, does not necessarily hold in the case of SBM. We must therefore test the validity of this assumption empirically.

In other words, we must provide evidence that the outcomes of students who met with mentors but did not form close relationships were not affected by mentoring. Assignment to the mentoring group, which is our instrument Z_i, may affect students who do not form close relationships if, say, SBM yields academic benefits as a result of the time those students spend working on homework and other academic activities with mentors. We thus propose and test a competing hypothesis to that which presumes the centrality of relationships: namely, that protégés' academic outcomes are affected by the time spent on academic activities in match meetings, whether or not the mentor-protégé relationship is a close one.[5]

The SBM programs in this data set differ by academic intensity. Although most devote little meeting time to academics, 225 of 565 treatment students were in programs that reported spending more than 25 percent of their meeting time on academic activities such as homework help and tutoring. We define such programs to be "academically focused," in contrast to the other, "relationship-only" programs. This variation in program focus endows the randomized evaluation with a cross-cutting design, in which two different treatments are tested simultaneously and in which there are three assignment groups (pure controls, students mentored in an SBM program with an academic focus, and students mentored in an SBM without such a focus). This design allows us to test the validity of the exclusion assumption in our instrumental

[5]Note that these two "competing" processes are not mutually exclusive. In fact, the presence of a close relationship between mentor and protégé may enable or intensify the effects of time spent on academic activities.

variable analyses. It also allows us to explore the effects of the interventions and their combination relative to the control group and relative to each other.

The following section reports the results of our empirical work. We first present evidence on the frequency and nature of close relationships in BBBSA SBM programs, as well as support for a simple characterization of relationships in this context. Next, we examine the evidence for academic benefits of SBM and report the effects of assignment to the mentoring treatment. Alongside the intention-to-treat estimates, we present various estimates of the effect of the treatment on the treated, using the approach outlined above, in order to determine if relationship closeness makes a difference. Third, we explore whether relationship quality still matters once relationship length and status are taken into account, and vice versa. We build upon the findings of previous analyses documenting the effects of match length and match status in school-based mentoring (Grossman et al., 2012) to refine the idea of "length plus strength" in SBM. Finally, we use logistic regression to investigate those features of SBM programs that may make close relationships more likely. Using the sample of all students assigned to treatment, we identify the program features and mentor characteristics associated with protégé reports of close mentoring relationships. We wish to provide guidance to SBM program administrators so that they can deliver the treatment of close relationships in more effective ways.

IV. Results

The Nature of Relationships in School-Based Mentoring

Many of the SBM matches were "successes," in that protégés characterized them as close. Of the 565 students assigned to treatment, 75.4 percent rated their relationships with their mentors as somewhat or very close at the time of the follow-up survey. The remaining 24.6 percent of treatment students were either never matched or rated their relationship as "not close at all" or "not very close."

Figure 1 reports the proportion of students who described their relationships with their mentors as somewhat or very close, by duration in months. A majority of matches of any duration are rated favorably by the protégés. After three months of meetings, over four-fifths of relationships are close, and that proportion does not increase appreciably with match duration — suggesting that most matches that ever achieved some degree of a close emotional connection did so within three months.

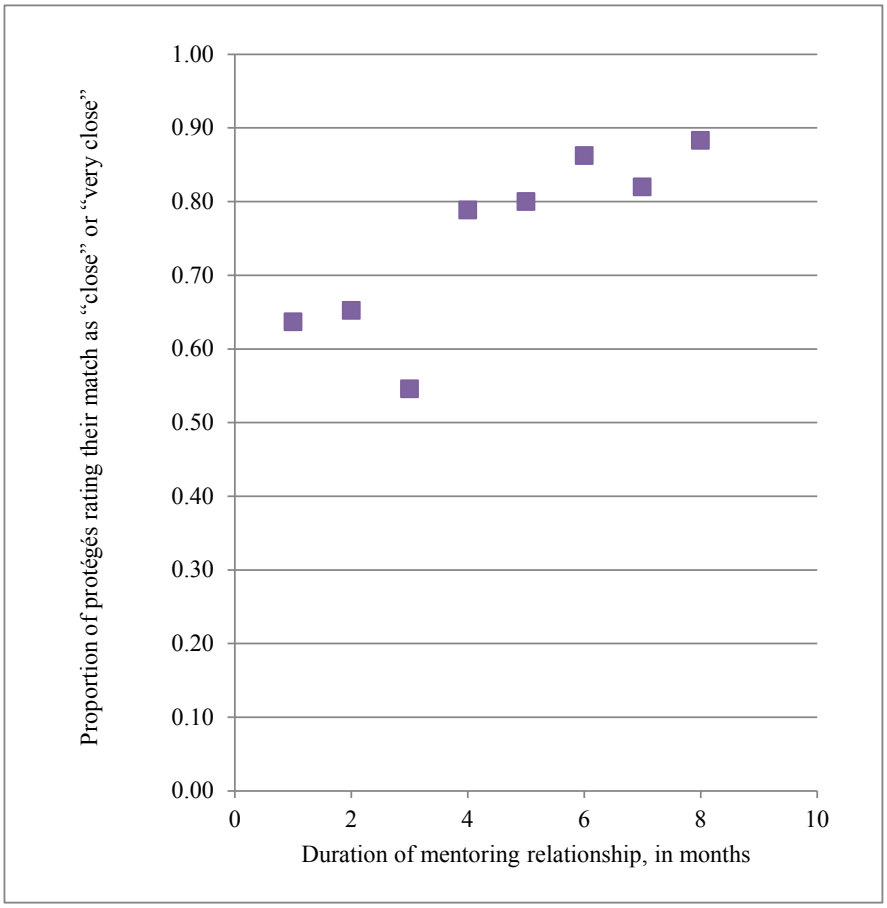
In addition, our initial analyses (not presented here; available upon request) suggest a threshold for closeness levels in SBM relationships. This threshold must be reached for a relationship to improve outcomes, but additional closeness beyond the threshold level has insignificant effects. We estimated the effect of relationship closeness on academic outcomes using detailed closeness ratings. Matches were categorized into four distinct groups: not close at all, not very close, somewhat close, and very close. We tested equality restrictions on the coefficients and found that mentoring experiences receiving either of the two lower closeness ratings produced statistically similar impacts, as did mentoring experiences receiving either of the two higher ratings (for overall academic performance, $F(2, 937) = 0.00$, $p = 0.9992$). These results support using a dichotomous measure of closeness in which relationships are recorded as "close" if and only if protégés rated them as "somewhat close" or "very close." The analyses reported in the following sections use this measure, modeling academic performance as a discontinuous function of closeness.

Impact of Relationship Quality on Academic Outcomes

Table 3 presents our main results on the academic benefits of school-based mentoring programs. It reports both the effects of assignment to the mentoring treatment and estimates of the role of relationships in SBM. The table reports impact coefficients and standard errors for each of the four academic outcome measures. Estimates reported as statistically significant in the table remain significant at a target level of significance of 0.05 after correcting for multiple comparisons (Benjamini & Hochberg, 1995).

Figure 1

Proportion of Matches with High Protégé Closeness Rating, by Duration in Months



Column (1) contains the OLS impact coefficients and standard errors from the intention-to-treat specification presented in equation (1). Consistent with the findings reported by Herrera et al. (2007), we find that students who were assigned to receive mentoring show statistically significant improvements in their academic performance and scholastic efficacy. Column (2) presents the potentially biased but straightforward treatment-on-treated estimates from OLS regression of outcomes on the treatment of interest, T_i, a close mentor-protégé relationship, as in equation (2). The estimates in column (2) are about 20 percent larger than those in column (1), as would be expected under the incomplete compliance with treatment. The

Table 3

Estimated Impact of a Close Relationship on Academic Outcomes

	(1) ITT	(2) TOT	(3) TOT controlling for relationship ability	(4) 2SLS
Overall academic performance (sample size = 945)				
Assigned to treatment	0.12**			
	(0.05)			
Received treatment		0.14***	0.14***	0.14**
(had a close relationship)		(0.05)	(0.05)	(0.06)
R^2	0.534	0.535	0.536	0.535
Quality of work (sample size = 953)				
Assigned to treatment	0.12***			
	(0.04)			
Received treatment		0.14***	0.14***	0.15***
(had a close relationship)		(0.04)	(0.04)	(0.05)
R^2	0.478	0.480	0.483	0.480
Completion of schoolwork (sample size = 953)				
Assigned to treatment	0.15***			
	(0.05)			
Received treatment		0.17***	0.17***	0.18***
(had a close relationship)		(0.05)	(0.05)	(0.06)
R^2	0.385	0.386	0.390	0.386
Scholastic efficacy (sample size = 1067)				
Assigned to treatment	0.07**			
	(0.03)			
Received treatment		0.08**	0.07**	0.08**
(had a close relationship)		(0.03)	(0.03)	(0.04)
R^2	0.268	0.269	0.275	0.269

NOTES: This table reports impact coefficients representing the estimated absolute change in the outcome measure associated with (assignment to) treatment. See Section III, "Outcome Measures and Key Variables," for an explanation of how outcome measures were derived. Column (1) presents the intention-to-treat (ITT) coefficients obtained from estimating equation (1). Column (2) presents the treatment-on-treated (TOT) coefficients obtained from estimating equation (2). Column (3) presents the treatment-on-treated (TOT) coefficients obtained from estimating equation (3), which includes a variable controlling for the quality of the protégé's relationship with his or her teacher. Column (4) presents the two-stage least squares (2SLS) estimates presented in equation (4). Standard errors are shown in parentheses. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent; and *** = 1 percent.

results indicate that a relationship rated as somewhat or very close by the protégé is associated with better academic outcomes.

In the next two columns, we attempt to learn more about causality. Does a close mentoring relationship *cause* the better academic outcomes? Our OLS estimates of the treatment-on-treated effect (column 2) may be biased due to the presence of a relevant but omitted variable. For instance, some children are better at working with adults, and these children may have both better mentoring relationships and better school performance.

Column (3), which corresponds to equation (3), addresses the potential omitted variable problem by including in our OLS specifications a variable that attempts to control directly for the possible confounding factor — namely, the quality of the child's relationship with his or her teacher before the mentoring program. This variable provides a measure of the child's ability to have a close relationship with a nonparental adult, and, as expected, has a positive and statistically significant effect on academic outcomes. As can be seen by comparing columns (2) and (3), including this teacher relationship variable does not affect our estimate of the impact of a close mentoring relationship.

Results when applying instrumental variable methods are shown in column (4). The estimated effects of a mentor-protégé connection in the 2SLS model are at least as large as the OLS estimates presented in column (2) and indicate that there are statistically significant academic benefits to a close mentoring relationship. We test to ensure that our instrument is not weakly correlated with the relationship variable after controlling for exogenous regressors, as weak instruments can produce estimates that are unreliable and biased toward the OLS estimate (Bound, Jaeger, & Baker, 1995). We find, as expected, that assignment to the mentoring group strongly predicts ($F = 1956$, $p = 0.00$) a connection with a program mentor. If we did have weak instruments, the 2SLS standard errors would be large and the validity of our instrument would be even more important, as even a slight correlation between error in the main equation and the instrument would lead to inconsistent estimates.[6]

Thus we find evidence that a close mentoring relationship positively affects academic performance. Effect sizes, obtained by dividing the impact coefficients reported in the table by the standard deviation of the appropriate outcome measure, range from 0.13 standard deviations (for overall academic performance and scholastic efficacy) to 0.18 standard deviations (for completeness of schoolwork), and are consistent across alternative specifications. The results presented in columns (3) and (4) suggest that there may not be an omitted variable bias issue in the OLS specification in column (2), removing the need for instrumental variable techniques in

[6]Tests of overidentifying restrictions for the model, although limited, do not suggest that the instruments are correlated with the error term in the main equation for any outcome.

follow-up analyses. The OLS and 2SLS estimates of the effect of close relationships are remarkably similar across specifications (2) through (4). A heteroskedasticity-robust Hausman (1978) test of endogeneity does not reject the null hypothesis that relationship closeness is exogenous, and the OLS specifications appear to provide unbiased estimates of the impact of providing a child with a close, caring relationship with a program mentor.

Table 4 provides further indication that these estimates can be interpreted as the causal effect of a close mentoring relationship. When we include a variable in our OLS regressions indicating the presence of a mentoring relationship that is "not close at all" or "not very close," we find no effect of mentoring for pairs that are not close. If the effects of "mentored but did not receive treatment (did not have a close relationship)" were statistically significant, we might conclude that SBM enhances student academic performance through alternative pathways. Instead, we find no evidence that mentoring enhances academic performance if the protégé does not have a good relationship with the mentor.

Table 4

Estimated Impact of Mentoring, With and Without a Close Relationship

	Overall academic performance	Quality of work	Completion of school-work	Scholastic efficacy
Received treatment (had a close relationship)	0.14*** (0.05)	0.14*** (0.05)	0.18*** (0.06)	0.07** (0.03)
Mentored but did not receive treatment (did not have a close relationship)	0.03 (0.09)	-0.04 (0.08)	0.05 (0.08)	-0.02 (0.06)
Sample size	945	953	953	1067
R^2	0.535	0.480	0.386	0.269

NOTES: This table reports impact coefficients representing estimated absolute changes in the outcome measures. See Section III, "Outcome Measures and Key Variables," for an explanation of how outcome measures were derived. Standard errors are shown in parentheses. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent; and *** = 1 percent.

We next use existing variation in program structure to explore explicitly the possibility that mentoring affects academic performance through additional time spent studying with a mentor rather than through the formation of a close mentor-protégé relationship. Some students were assigned mentors in relationship-only programs, whereas others were mentored in academically focused programs. In each type of program, some students connected with their assigned mentors and thus received treatment, and others did not. Relationship-formation rates are essentially identical across program types. Of the 225 students assigned to treatment in academically focused programs, 75.1 percent connected with their mentors by the time of the

follow up survey; in relationship-only programs, 75.6 percent of protégés reported feeling close to a mentor.

The specifications in Table 5 are like those in columns (1) and (4) of Table 3 but they also include an indicator variable specifying whether or not a student was at a site that ran an academically focused program, and two treatment dummies: one indicating students assigned to the treatment group at a relationship-only site and one indicating students assigned to treatment at an academically focused site. In columns (1) and (3), we constrain the impacts across these two types of sites to be the same, whereas in columns (2) and (4) we allow them to differ and test whether the constraint is valid.

The first two columns of Table 5 present intention-to-treat estimates for the constrained and unconstrained specifications. In general, assignment to the treatment group appears to have similar effects whether the mentoring occurs with an academic focus or not. For overall academic performance, the effects of the two different treatment conditions are virtually identical. In completion of work students mentored in an academically focused program perform worse, while in scholastic efficacy they perform better. Indeed, the effects are not significantly different across program types for all four outcome measures.[7] Thus we see no evidence that students mentored in an academically focused program do better academically than students in relationship-only programs.

We also delve more deeply to ascertain the differences in treatment effects on the treated, though we do not expect a different verdict, given the similar rates of relationship formation in each program type. Columns (3) and (4) of Table 5 present the 2SLS treatment-on-treated estimates for the constrained and unconstrained specifications, respectively. Here we use an expanded instrument set created by interacting treatment assignment with program type, which allows us to disentangle the effects of relationship and direct academic help since the two instruments may have differential effects on these pathways. Column (3) uses one endogenous regressor and thus is overidentified. Column (4) has two endogenous regressors (our relationship closeness variable in each of the two program types) and thus is just-identified.

We find that an academic focus does not produce a detectably different effect of having a close SBM relationship. Following Angrist (1991), we use the overidentification test statistic in column (3) to test whether the 2SLS estimates of the effects of the two program types are equal in column (4). We find they are not significantly different (all p-values are greater than

[7]Note that the data may not have sufficient power to distinguish the effects of the programs. Particularly in the cases of completion of schoolwork and of student-reported scholastic efficacy, failure to reject equality may be a type II error. Both also fail to reject a null hypothesis of a quarter standard deviation difference in effects, but the effects are in opposite directions.

Table 5

Estimated Impact on Academic Outcomes, by Program Type

	(1) ITT	(2) ITT	(3) 2SLS	(4) 2SLS
Overall academic performance				
Assigned to treatment	0.12**			
	(0.05)			
…in an RO program		0.12*		
		(0.06)		
…in an AF program		0.12		
		(0.08)		
Received treatment (close relationship)			0.15**	
			(0.06)	
…in an RO program				0.15*
				(0.08)
…in an AF program				0.15
				(0.10)
Overidentification test $\chi^2(1)$			0.000	
			(p = 0.990)	
Quality of work				
Assigned to treatment	0.12***			
	(0.04)			
…in an RO program		0.10*		
		(0.05)		
…in an AF program		0.15*		
		(0.08)		
Received treatment (close relationship)			0.15***	
			(0.05)	
…in an RO program				0.12*
				(0.06)
…in an AF program				0.18*
				(0.10)
Overidentification test $\chi^2(1)$			0.260	
			(p = 0.610)	
Completion of schoolwork				
Assigned to treatment	0.15***			
	(0.05)			
…in an RO program		0.18***		
		(0.07)		
…in an AF program		0.10		
		(0.09)		
Received treatment (close relationship)			0.18***	
			(0.06)	
…in an RO program				0.22***
				(0.08)
…in an AF program				0.12
				(0.11)
Overidentification test $\chi^2(1)$			0.441	
			(p = 0.507)	

(continued)

Table 5 (continued)

Scholastic efficacy				
Assigned to treatment	0.07**			
	(0.03)			
…in an RO program		0.03		
		(0.04)		
…in an AF program		0.14**		
		(0.06)		
Received treatment (close relationship)			0.09**	
			(0.04)	
…in an RO program				0.04
				(0.05)
…in an AF program				0.17**
				(0.07)
Overidentification test $\chi^2(1)$			2.39	
			(p = 0.122)	

NOTES: This table reports impact coefficients representing the estimated absolute change in the outcome measure associated with (assignment to) treatment. See Section III, "Outcome Measures and Key Variables," for an explanation of how outcome measures were derived.

AF = academically focused, RO = relationship-only, ITT = intention-to-treat, and 2SLS = two-stage least squares.

Columns (1) and (3) constrain the impacts across AF and RO sites to be the same, whereas in columns (2) and (4) we allow them to differ.

The overidentification test statistic does not reject the hypothesis that effects of the two program types are equal (Angrist, 1991).

Standard errors are shown in parentheses. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent; and *** = 1 percent.

0.10), and thus we find no evidence of a causal link between assigned treatment status and the outcomes other than the close relationship formed between mentor and protégé.

Relationship Quality vs. Relationship Length and Status

In this section, we collate our findings with those of previous studies on the effects of mentoring match length and status. Innovations developed in this paper clarify previously published results about effective mentoring practice.

In Table 6, we first reproduce the apparent benefits of intactness, originally reported by Grossman et al. (2012), for each outcome measure. When we interact the match closeness indicator with match status, however, we find that only close, intact matches have an impact. Protégés in intact matches who are not close to their mentors fared no better than they would have without mentors. Protégés who were rematched and had a close relationship with their

Table 6

Estimated Impact of Match Status, Controlling for Relationship Closeness

	Overall academic performance		Quality of work		Completion of schoolwork		Scholastic efficacy	
Intact match	0.11**		0.11**		0.15***		0.07**	
	(0.05)		(0.05)		(0.05)		(0.03)	
Rematch	0.08		0.14		0.15		0.03	
	(0.10)		(0.09)		(0.12)		(0.07)	
No rematch	0.06		0.14		0.01		0.09	
	(0.14)		(0.10)		(0.15)		(0.08)	
Not close/intact match		-0.01		-0.08		0.07		-0.02
		(0.11)		(0.10)		(0.10)		(0.07)
Not close/rematch		0.03		-0.08		-0.11		-0.06
		(0.27)		(0.26)		(0.23)		(0.11)
Not close/no rematch		-0.03		0.16		0.03		0.15
		(0.21)		(0.16)		(0.22)		(0.15)
Close/intact match		0.13**		0.12**		0.17***		0.08**
		(0.06)		(0.05)		(0.06)		(0.04)
Close/rematch		0.09		0.18**		0.21*		0.04
		(0.10)		(0.09)		(0.13)		(0.08)
Close/no rematch		0.24		0.21		0.17		0.02
		(0.20)		(0.14)		(0.24)		(0.09)
R^2	0.533	0.534	0.478	0.480	0.385	0.387	0.268	0.270

NOTES: This table reports impact coefficients representing estimated absolute changes in the outcome measures. See Section III, "Outcome Measures and Key Variables," for an explanation of how outcome measures were derived. Standard errors are shown in parentheses. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent; and *** = 1 percent.

new mentors were not negatively affected. Indeed, there is evidence that rematched students in close relationships were doing significantly better than their control peers with respect to the quality of their schoolwork, and it is possible that the positive impacts of being matched had not yet grown large enough to be statistically significant for the rematched students' other outcomes. Thus, the current study refines the work of Grossman et al. (2012) to suggest that rematching does not have negative effects as long as the new relationship clicks.

In Table 7, we first provide analyses that include indicators of match duration but not closeness and reproduce the stylized fact that longer matches yield greater benefits. Once we control for match closeness by interacting the match closeness indicator with match duration, however, we observe that only close relationships have impacts and that match length does not matter much. In other words, if a match is long but not good, it is not productive; if it is short but good, it is productive.

Our results — that longer matches do not necessarily produce greater benefits in SBM — are therefore consistent with Grossman et al. (2012). Since Tables 6 and 7 report the results of OLS regressions, we take steps to address possible omitted variable bias even though the results of the previous section suggest such steps are unnecessary. While we do not have an adequate number of instruments to produce 2SLS estimates, when we add the teacher relationship variable to the specifications to control for variation in student relationship ability, we find results similar to those reported in the tables.

Facilitating Relationships in School-Based Mentoring

The evidence compiled in the previous sections suggests that a close relationship between mentor and protégé may be the key to the effectiveness of SBM. This section explores the characteristics of programs and mentors that may increase the likelihood of a close relationship. The goal is to help administrators learn how better to design SBM programs so as to facilitate relationship formation.

Table 8 reports evidence on the program features and mentor characteristics associated with close mentor-protégé relationships. Using the sample of all students assigned to treatment, we employ logistic regression to identify the factors influencing the likelihood that a protégé felt close to his or her mentor, enriching the findings reported by Herrera et al. (2007). Column (1) presents odds ratios from our main specification. We found several statistically significant associations, but program type (academically focused or relationship-only) was not one of them. A program with an academic focus did not help or hinder relationship formation. Other program features appeared to be very consequential: protégés less frequently reported close connection with their mentors in programs where all pairs met at the same time and location and in those where pairs met only once or twice a month. Column (2) further investigates the negative effect

Table 7

Estimated Impact of Match Length, Controlling for Relationship Closeness

	Overall academic performance		Quality of work		Completion of schoolwork		Scholastic efficacy	
Short	0.00		0.06		0.13		0.04	
	(0.09)		(0.08)		(0.11)		(0.06)	
Medium	0.13*		0.13**		0.14*		0.06	
	(0.07)		(0.06)		(0.08)		(0.05)	
Long	0.14**		0.13**		0.16***		0.09**	
	(0.06)		(0.05)		(0.06)		(0.04)	
Not close/short		-0.08		-0.10		-0.03		0.03
		(0.19)		(0.17)		(0.17)		(0.10)
Not close/medium		0.14		0.04		0.14		0.00
		(0.13)		(0.13)		(0.16)		(0.10)
Not close/long		0.01		-0.00		0.04		-0.05
		(0.14)		(0.12)		(0.10)		(0.09)
Close/short		0.06		0.15*		0.26*		0.04
		(0.10)		(0.09)		(0.14)		(0.08)
Close/medium		0.11*		0.12*		0.15*		0.08
		(0.07)		(0.07)		(0.09)		(0.05)
Close/long		0.19***		0.15***		0.18***		0.10**
		(0.08)		(0.06)		(0.06)		(0.04)
R^2	0.534	0.537	0.478	0.481	0.385	0.387	0.269	0.270

NOTES: This table reports impact coefficients representing estimated absolute changes in the outcome measures. See Section III, "Outcome Measures and Key Variables," for an explanation of how outcome measures were derived. Standard errors are shown in parentheses. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent; and *** = 1 percent.

of all matches meeting at the same time and location and finds that the large-group setting impaired relationship formation whether or not the mentor was a high school student.

Protégés were less likely to feel close to college student mentors than to adult volunteers, but more likely to feel close to high school student mentors. Same-race and same-gender matches were not more likely to be close. As column (3) shows, mentors who showed up when they were supposed to, as rated by the protégés, were significantly more likely to be in close relationships, especially when the program's meetings were infrequent. (The possibility of endogeneity bias does exist though, as a mentor may not have shown up if the pair was not feeling particularly close.) Lastly, mentors who received more training before and during their mentoring assignments were more likely to have close relationships with their protégés.

Although we have data on several other program features discussed in the literature (such as parent involvement and the number of contacts between the mentor and the case manager), we did not estimate the associations with variables not directly controlled by program administrators or those we believed were likely to be too endogenous. The factors shown in Table 8 are ones that programs can control and that may make it more likely for pairs to form close relationships.

Table 8

Predictors of a Close Relationship (Odds Ratios)

	(1)	(2)	(3)
Program features			
Academic focus	0.69	0.70	0.24
	(0.47)	(0.46)	(0.23)
All matches meeting at same time and location	0.31***		0.20***
	(0.13)		(0.12)
if high school student mentor		0.24***	
		(0.13)	
if not high school student mentor		0.40**	
		(0.18)	
Some structured time	1.35	1.33	1.25
	(0.59)	(0.58)	(0.83)
Infrequent meetings	0.28***	0.29***	0.00***
	(0.13)	(0.13)	(0.00)
Volunteer recognition	1.01	1.04	1.98*
	(0.15)	(0.16)	(0.70)
Age of student	0.82***	0.82***	0.71***
	(0.05)	(0.05)	(0.09)
Mentor characteristics			
High school student	1.28	1.66	1.96
	(0.45)	(0.74)	(1.11)
College student	0.27***	0.27***	0.47
	(0.12)	(0.12)	(0.46)
Same gender as student	0.88	0.88	1.82
	(0.37)	(0.37)	(1.24)
Same race as student	1.15	1.14	1.39
	(0.33)	(0.33)	(0.79)
Mentor shows up			1.80***
			(0.30)
Mentor shows up/infrequent program meetings			3.17**
			(1.71)
Amount of training mentor received			1.34***
			(0.12)
Sample size	474	474	271

NOTES: This table presents the odds ratios associated with various program and mentor characteristics. Column (1) is our main specification. Column (2) allows the effect of the matches meeting at the same time and location to vary across mentor student status. Column (3) includes additional mentor characteristics. Specifications include site fixed effects and youth characteristics (for example, gender, race, or eligibility for free lunch). Cluster-robust standard errors are shown in parentheses. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent; and *** = 1 percent.

V. Summary and Discussion

This study finds evidence that unless a mentorship pair forms a relationship that the protégé would rate as "somewhat close" or better, SBM program participation has no effect on a student's end-of-school-year academic outcomes. We thus posit that this close relationship is the "active ingredient" of mentoring in this setting. This conclusion is very much in line with much of the theoretical work done on mentoring (Rhodes, 2005) and suggests that BBBSA's school-based mentoring does not work merely as a tutoring program or by providing a context for specific training, as Cavell and Elledge suggest (forthcoming). We also find that SBM programs that focused more heavily on academic activity had no larger academic effects than those where mentors engaged primarily in social activities.

Much of the field's thinking about mentoring developed from community-based mentoring programs. That common wisdom says it takes time, often months, for a relationship to "gel," for trust to be built so that the match can affect the life of a child (Grossman & Rhodes, 2002). This study challenges this assumption for mentoring in the school setting by showing that a majority of protégés reported fairly close relationships even if they had been matched only for a month or two, and that such relationships accounted for positive impacts on academic outcomes. Perhaps the school setting encourages children to trust these unrelated adults more quickly, or children may have lower thresholds for close relationships with "school adults" than with "community adults." The fact that relationships between protégés and mentors in the BBBSA SBM program seem to gel fairly quickly may be why in their analyses with these same data Grossman et al. (2012) find little relationship between months of mentoring and academic impacts.

Program staff and empirical researchers alike often worry that the children able to form the best relationships are those who would have normally done well anyway — in school or life — because of hard-to-measure characteristics such as persistence, or a child's ability to get along with adults. But we find little evidence for this type of sorting among protégés with close relationships. If this type of process were occurring, we would see large differences between the 2SLS impact estimates (which account for this potential self-selection) and the estimates obtained from the simpler OLS regressions. From a technical point of view, this finding is good news because it means we can control for closeness directly, without having to worry that this indicator is correlated with the error term. But from a programmatic point of view the finding is even more important, as it suggests that BBBSA SBM programs can be similarly helpful to all types of students.

Our estimated treatment-on-treated effect sizes, ranging from 0.13 to 0.18 standard deviations, compare favorably to the changes in the outcome measures for the control students

over the course of the year. For control students, overall academic performance improved by 0.054 (standard deviation = 0.814) and quality of work by 0.016 (standard deviation = 0.729), while completion of work decreased by 0.038 (standard deviation = 0.873) and scholastic efficacy by 0.034 (standard deviation = 0.612). Moreover, the estimated effects of a close relationship with a mentor assigned through the BBBSA SBM program are comparable to the size of the impacts of many other school-based interventions on academic outcomes. For example, in Project STAR, the academic achievement impacts of reduced class sizes were an estimated 0.1 to 0.2 standard deviations (Krueger, 1999). Similarly, a meta-analysis of after-school programs reported average effect sizes ranging from 0.05 standard deviations on reading to 0.17 on math skills (Lauer, 2006). Another meta-analysis of explicitly nonacademic after-school programs (Durlak & Weissberg, 2007) found impacts on grades ranging from 0.24 standard deviations for programs using evidence-based approaches to 0.05 for those not incorporating best practices.

This study also sheds light on the perplexing rematch result reported by Grossman et al. (2012) in analyses of the same data analyzed in this paper. That study found that even though rematched protégés had slightly more total weeks of mentoring than protégés in intact matches, they performed *less* well than students who never had a mentor. The authors had no explanation for the result, but hypothesized that perhaps the relationships had not gelled sufficiently. That explanation does not seem plausible now that we know how quickly school-based mentoring relationships can click, at least from the protégé's perspective.

Indeed, once we control for the quality of the matches, we see that although close, intact matches have the largest — and statistically most significant — estimated impacts on academic outcomes, rematched students in close relationships do not appear to be hurt. In all cases the estimated impacts are positive and fairly similar to those of the students who remained in their original mentoring relationships, though the standard errors on these coefficients are large. The negative effects appear to be centered among the rematched protégés who do not feel close to their new mentors (though because the subsamples are small, these negative coefficients are not statistically significant). These findings indicate that the emotional tenor of the relationship in the new match is the key issue programs should address. It could be that the second matches were rushed and less well suited to the students than their original matches. It could be that protégés felt rejected when their first mentoring relationships ended, affecting their ability to feel close to another program-assigned mentor quickly. More research should be done to investigate these issues.

Similarly, this paper investigates the "length plus strength" hypothesis that grew out of earlier work on the BBBSA CBM program, in which estimated impacts were observed to increase over time once it appeared a trusting relationship had formed (Grossman & Rhodes, 2002). We find that only close relationships have impacts and that match length is not as im-

portant in the SBM programs we study. We do not use the measure of length that has been used in other match-length studies (the total number of weeks a protégé meets with any mentor), as we are interested in understanding whether the impact a particular mentor has on a protégé increases with the length of their relationship. We do find evidence that long and close relationships led to improvements in academic outcomes, but it is difficult to conclude definitively that shorter but close relationships had smaller or no positive effects. Positive impacts on the completion and quality of the student's homework appear to happen rather quickly, while improvements in the teacher's overall assessment of the student's performance and in the student's assessment of his or her own scholastic competence do appear to grow larger with longer relationships. More research involving large numbers of students is needed to answer this question.

The analyses carried out in this study suggest that the BBBSA SBM program (and potentially others with similar characteristics) can make close relationships more likely by incorporating several program practices:

- Not having pairs meet in the same area at the same time
- Giving mentors structured activities to do with their protégés
- Asking mentors to come at least three times a month
- Providing BBBSA staff support to mentors

It might be a challenge for underfunded schools to provide adequate mentoring space, especially if mentors and protégés need multiple locations to meet on their own: schools might have to pay custodians more to clean the spaces after hours, for example. Schools might be able to reduce costs by encouraging pairs to meet during students' lunch, recess, or study hall time.

Although we did not find that academically focused mentoring led to better academic outcomes if the mentoring relationship was not close, it still may be justified to center well-designed volunteer programs on academic help. The mentors studied here were given no instruction in tutoring, nor did they have a structured curriculum to follow. The evidence from this study should therefore not be used to conclude that volunteer tutoring programs are ineffective. But greater attention to the importance of closeness could help inform the design of tutoring programs and other after-school programs. Program designers would do well to incorporate practices that make it easier for staff members and volunteers to form close relationships with the young participants and thereby potentially strengthen program effects.

This study develops an empirical framework to investigate rigorously the mechanisms by which a mentoring program may improve academic outcomes. Given that we find the mentor-protégé relationship to be the core element of mentoring, we believe that any analysis of experimental data on mentoring programs should consider the partial compliance model used in

this paper. Although instrumental variables and other nonexperimental techniques do not produce results as convincing as evidence from randomized control trials, in the end it is not possible to randomly assign close relationships. Statistical techniques can appropriately adjust for the presence or absence of closeness. Building on the results of this study, future experiments can include designs created to refine our understanding of the mechanism behind mentoring effectiveness, with randomly selected mentoring matches or programs encouraged to incorporate either more relationship-building activities or more academically focused ones.

Many schools are struggling to meet the standards set by the No Child Left Behind Act of 2001, and the future holds even more mandates and struggles as states implement the Common Core Standards. Under binding budget constraints, volunteer labor will be increasingly attractive. We find that school-based mentoring programs that aim to provide relationships can improve academic outcomes. Not every adult can be a good tutor, but many of them could help students academically by being a caring presence.

References

Angrist, J.D. (1991). Grouped-data estimation and testing in simple labor-supply models. *Journal of Econometrics*, 47(2-3), 243-66.

Angrist, J.D., Imbens, G.W., & Rubin, D.B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434), 444-455.

Benjamini, Y., & Hochberg, Y. (1995). Controlling for the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, 57(1), 289-300.

Bernstein, L., Dun Rappaport, C., Olsho, L., Hunt, D., & Levin, M. (2009). *Impact evaluation of the U.S. Department of Education's Student Mentoring Program* (NCEE 2009-4047). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Big Brothers Big Sisters of America. (2013a). Our programs get things started. Retrieved from: http://www.bbbs.org/site/c.9iILI3NGKhK6F/b.5968193/k.5031/Our_programs_get_things_started.htm.

Big Brothers Big Sisters of America. (2013b). School-based programs. Retrieved from: http://www.bbbs.org/site/c.9iILI3NGKhK6F/b.5961387/k.6048/SchoolBased_Programs.htm.

Big Brothers Big Sisters of America. (2013c). We are here to start something. Retrieved from: http://www.bbbs.org/site/c.9iILI3NGKhK6F/b.5962351/k.42EB/We_are_here_to_start_something.htm.

Bloom, H. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review*, 8(2), 225-246.

Bound, J., Jaeger, A., & Baker, R. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90, 443-450.

Cavell, T.A., & Elledge, L.C. (Forthcoming). Mentoring and prevention science. In D.L. DuBois & M.J. Karcher (Eds.), *Handbook of Youth Mentoring* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Cronbach, L.J. (1951). "Coefficient alpha and the internal structure of tests." *Psychometrika*, 16, 297-334.

DuBois, D.L., & Neville, H.A. (1997). Youth mentoring: Investigation of relationship characteristics and perceived benefits. *Journal of Community Psychology*, 25, 227-234.

Durlak, J.A., & Weissberg, R.P. (2007). *The impact of after-school programs that promote personal and social skills*. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning.

Durlak, J.A., Weissberg, R.P., Dymnicki, A.B., Taylor, R.D., & Schellinger, K.B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82(1), 405-432.

Grossman, J.B., Chan, C.S., Schwartz, S.E., & Rhodes, J.E. (2012). The test of time in school-based mentoring: The role of relationship duration and re-matching on academic outcomes. *American Journal of Community Psychology*, 49(1-2), 43-54.

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, 199-219.

Harter, S. (1985). *Manual for the self-perception profile for children.* Denver, CO: University of Denver.

Harter, S. (1990). Identity and self development. In S. Feldman and G. Elliott (Eds.), *At the threshold: The developing adolescent.* Cambridge, MA: Harvard University Press.

Hausman, J.A. (1978). Specification tests in econometrics. *Econometrica*, 46(6), 1251-1271.

Herrera, C. (2004). *School-based mentoring: A closer look*. Philadelphia, PA: Public/Private Ventures.

Herrera, C., DuBois, D.L., & Grossman, J.B. (2013). *The role of risk: Mentoring experiences and outcomes for youth with varying risk profiles*. New York, NY: A Public/Private Ventures project published by MDRC.

Herrera, C., Grossman, J.B., Kauh, T.J., Feldman, A.F., & McMaken, J. (2007). *Making a difference in schools: The Big Brothers Big Sisters school-based mentoring impact study.* Philadelphia, PA: Public/Private Ventures.

Imbens, G., & Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62, 467-75.

Institute of Education Sciences. (2011). *WWC procedures and standards handbook* (Version 2.1). Washington, DC: Institute of Education Sciences, U.S. Department of Education.

Karcher, M.J. (2003). The Hemingway: Measure of adolescent connectedness. Retrieved from http://www.adolescentconnectedness.com.

Krueger, A.B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, 114(2), 497-532.

Lauer, P.A., Akiba, M., Wilkerson, S.B., Apthorp, H.S., Snow, D., & Martin-Glenn, M.L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, 76, 275-313.

MENTOR. (2006). *Mentoring in America 2005: A snapshot of the current state of mentoring*. Retrieved from http://www.mentoring.org/downloads/mentoring_333.pdf.

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

Pierce, K.M., Hamm, J.V., & Vandell, D.L. (1999). Experiences in after-school programs and children's adjustment in first-grade classrooms. *Child Development,* 70(3), 756-767.

Pryce, J., & Keller, T.E. (2012). An investigation of volunteer-student relationship trajectories within school-based youth mentoring programs. *Journal of Community Psychology*, 40(2), 228-248.

Rhodes, J.E. (2005). A model of youth mentoring. In D.L. DuBois, & M.J. Karcher (Eds.), *Handbook of youth mentoring* (pp. 30-43). Thousand Oaks, CA: Sage Publications.

Thomson, N.R., & Zand, D.H. (2010). Mentees' perceptions of their interpersonal relationships: The role of the mentor-youth bond. *Youth & Society*, 41(3), 434-445.

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

Zand, D., Thomson, N.R., Cervantes, R., Espiritu, R., Klagholz, D., LaBlanc, L., & Taylor, A. (2009). The mentor-youth alliance: the role of mentoring relationships in promoting youth competence. *Journal of Adolescence*, 32, 1-17.