

ARTICLE

Mentoring program practices as predictors of match longevity

Janis B. Kupersmidt¹ | Kathryn N. Stump¹ | Rebecca L. Stelter¹ |
Jean E. Rhodes²

¹Innovation Research & Training

²University of Massachusetts Boston

Correspondence

Email: jkupersmidt@irtinc.us

We thank Big Brothers Big Sisters of America (BBBSA) for participating in and providing access to the archive of data analyzed in this study, and to Greg Zweber and Dr. Jessica Mitchell from BBBSA for helpful feedback on this paper.

This project was supported by Grant #2012-JU-FX-0009 awarded to the first author by the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice.

Abstract

Mentoring has been shown to have a small to moderate effect on youth outcomes; however, implementation of research-based practices improves program efficacy. Benchmark program practices and Standards in the Elements of Effective Practice for Mentoring (MENTOR, 2009) were assessed in the current study as predictors of match longevity. Secondary data analyses were conducted on a national agency information management database from 45 Big Brothers Big Sisters agencies and 29,708 matches from across the U.S. Agency leaders completed a web-based survey describing their policies, procedures, and practices. Results revealed that the only individual Standard that was predictive of match length was the Training Standard. In addition, the sum total of both Benchmark program practices and Standards were associated with match length and long-term relationships; however, neither predicted premature match closure. Results are discussed in terms of the importance of implementation of evidence-based practices, particularly mentor training, for achieving effective mentoring outcomes.

1 | INTRODUCTION

Mentoring relationships are associated with positive academic, social, emotional, and behavioral outcomes for youth; however, despite consistently positive findings across studies, the overall effect size for mentoring is fairly modest (DuBois, Portillo, Rhodes, Silverthorn, & Valentine, 2011). In addition, across outcome studies, the effect sizes range considerably, suggesting that some programs have a greater impact on youth than others. Positive youth development programs such as mentoring programs have become increasingly more accountable for reporting on the effectiveness of their interventions, with attention focused not only on outcomes for participating youth, but also on the quality of the programs themselves (Yohalem & Wilson-Ahlstrom, 2010). This emerging interest in program operations and practices extends beyond a simple audit or process evaluation of whether or not programs are implementing their practices as they were designed. There is also ample reason to examine program operations in the form of program practices. Following this trend, research is needed on understanding which mentoring program practices, if any, may enhance match and youth outcomes.

1.1 | Elements of effective practice for mentoring

Drawing on the growing body of research on mentoring, the groundbreaking manual, *Elements of Effective Practice for Mentoring* (EPEM; MENTOR, 2009), was written by a small team of researchers (Kupersmidt, Stelter, and Rhodes) and practitioners, after conducting a thorough review of the empirical and practice literatures on mentoring. The review focuses on identifying program practices that were designed to protect the safety of mentees as well as empirically associated with positive match or youth outcomes. Thus, the EPEM contains practices that are proximal to the development and maintenance of an effective mentoring relationship, and are central to protecting the physical and emotional safety of mentees. Consequently, the EPEM contains the authoritative operational standards and benchmark practices for designing effective mentoring programs.

The EPEM is organized around six standards that follow the life cycle of a mentoring relationship, beginning with recruitment of participants into a mentoring program through closure of the mentoring relationship, and comprises a total of 23 benchmark practices distributed across these standards. In order for a mentoring program to meet a standard, the program must implement all the benchmarks associated with the standard. The EPEM is agnostic with respect to gradations in the value or degree of importance of each benchmark because all of the benchmarks are considered fundamental to effective programming. However, the validity of this hypothesis has never been tested. In other words, it is not yet known whether the individual standards are of equal importance for achieving positive outcomes, nor whether the comprehensive implementation of the benchmark practices is actually necessary to achieve more positive outcomes than less adherence to the full set of benchmark practices. Thus, exploring the relationship between program standards and benchmark practice implementation and match outcomes, particularly those related to match longevity, forms the main goal of the present study.

1.2 | Program practices and outcomes

There are few large-scale studies that provide descriptive information on the current use of research-based practices by mentoring programs nationally and even fewer that provide information about the prevalence of usage of these practices in programs serving high-risk youth. For example, despite the fact that mentor training and support practices directly affect mentee-related outcomes (DuBois, Holloway, Valentine, & Cooper, 2002), the use of these practices is far from universal. In a survey of a diverse set of over 700 mentoring programs, most (more than 90%) reported providing their volunteers with some orientation or training; however, 14% required only an orientation, but no training (Sipe & Roder, 1999). In a meta-analysis of evaluations of 53 researched mentoring programs, prematch training or orientation was provided to mentors in only 71% of the programs and ongoing training was provided to postmatch mentors in only 23% of the programs (DuBois et al., 2002).

In addition, 131 Big Brothers Big Sisters (BBBS) agencies were surveyed 83% of the programs reported that they provided prematch orientation/training of varying lengths to volunteers. (Wheeler & DuBois, 2009). Prematch mentor training and postmatch mentor support are common, but are not universally implemented across practice settings, suggesting substantial room for quality improvement in program practices. Furthermore, of the studies that have been located, none comprehensively examined all of the benchmark practices in the EPEM. Thus, a second goal of the present study was to examine the prevalence of implementation of the EPEM benchmark practices within a subset of agencies using the BBBS program model.

Match duration is an important outcome to study because it serves as a moderator of youth outcomes for mentoring relationships and the minimum recommended match duration of one calendar (for community-based mentoring) or school (for site-based mentoring) year is difficult to achieve. Longer term mentoring relationships have been found to be associated with more benefits to youth than shorter term relationships (MENTOR, 2015). As match length increases, positive outcomes for youth also increases, with the greatest benefits occurring for matches that last one year or longer (Grossman & Rhodes, 2002; Grossman, Chan, Schwartz, & Rhodes, 2011).

Also, match length has been positively associated with match quality, as rated by mentors and mentees (Rhodes, Schwartz, Willis, & Wu, 2014). Unfortunately, matches that terminate prematurely, prior to one year of length, are

associated with negative outcomes for youth (Herrera et al., 2007). Despite the advantages associated with promoting and supporting match longevity, premature match closure is unfortunately too common. In several national studies of mentoring programs, only 43% of school-based matches lasted 6 months or longer (Bernstein, Rappaport, Olsho, Hunt, & Levin, 2009) and 38% of matches from a wide range of program models ended prematurely (Kupersmidt, Stump, Stelter, & Rhodes, in press). Because of the need to understand factors that promote match longevity and prevent premature match closure, program practice implementation was examined in this study in relation to these indicators of match effectiveness.

1.3 | The current study

The current study examines archival data on match longevity outcomes from the BBBS of America's (BBBSA) national data archive and program practices data from web-based questionnaires completed by program leaders at local BBBS agencies. The main goal was to explore the prevalence of implementation of program practice benchmarks and standards, as well as the consequences of implementation fidelity on match length. Consistent with previous research and theory on moderators of match outcomes, we hypothesized that greater compliance with evidence- and safety-based program practices would be associated with more success in fostering longer matches.

2 | METHOD

2.1 | Participants

The participants in this study included 45 BBBS local agencies located in 28 states in the United States who used the BBBSA Agency Information Management (AIM) software for managing their programs. Of the 194 BBBS programs invited to participate, 23% agreed to participate, provided informed consent, and participated in the project. Programs that agreed to participate in the project did not differ from nonparticipating programs in terms of their average match length (participating programs = 19.1 months; nonparticipating programs = 19.2 months; $t_{(192)} = 0.08$, *ns*), or their total number of community-based matches (participating programs = 660 matches; nonparticipating programs = 351 matches; $t_{(47.8)} = 1.84$, $p = 0.07$).

Participating programs ranged in size, cumulatively, from having fewer than 30 matches represented in the dataset to more than 13,000. All matches were one-to-one relationships; approximately 50% of the matches were community-based, and 50% of the matches were school- or site-based. For purposes of the current project, all analyses were conducted using community-based matches, thus, yielding a final sample size of 29,708 matches. Data from the matches were entered into the AIM data management system from October 1, 2004, through March 30, 2013.

Most of the program staff who completed the surveys were women (91%), White (82%), in their 40s (mean age of 42, range from 24 to 65), and had been at their agency for an average of 13.5 years (range from 1.5 to 40 years). The majority of mentors were women (63%), White (66%), and non-Hispanic (91%), with an average age of 31.74 years (standard deviation [*SD*] = 11.99). The majority of mentees were girls (59%). In contrast to the demographic composition of the group of mentors, mentees were more balanced in terms of race; 28% of the mentees were White and 17% were Hispanic. Mentees were, on average, 10.78 years old ($SD = 2.47$) at the beginning of their match.

2.2 | Measures

2.2.1 | Archival AIM data

2.2.1.1 | Demographic characteristics

Basic demographic information included sex, race, ethnicity, and age of mentors and mentees. Race and ethnicity were coded as two dichotomous variables indicating White versus non-White mentees and non-Hispanic versus Hispanic mentees, respectively. BBBS staff collected data when mentors and mentees applied to participate in the program.

Because data were collected over the span of 10 years, a four-category cohort variable was constructed for the purpose of including it as a covariate in the main analyses. The sample was divided into four cohorts to have an approximately equivalent number of matches in each cohort. Matches that began before 2009 were placed in cohort 1 ($n = 6,254$); matches that began in 2009 were placed in cohort 2 ($n = 7,185$); matches that began in 2010 were placed in cohort 3 ($n = 7,988$); and matches that began in 2011 or later were placed in cohort 4 ($n = 10,491$), and served as the referent group.

2.2.1.2 | Match relationship length

Relationship length information was calculated based upon agency records of the start and end dates of each match. If matches were still open on the date when the study ended (i.e., the date that the data were downloaded from BBBSA's AIM database), then match length was calculated as the time between the first date of the match and the last day of the study. In addition, a dichotomous variable was created to indicate whether the match was open or closed.

2.2.1.3 | Premature match closure

BBBS agencies request a minimum 12-month commitment from their mentors, when they participate in their community-based mentoring programs. Therefore, matches with lengths less than 12 months were coded as prematurely ending. To avoid coding open matches that were recently made as matches that ended prematurely, open matches that were less than 12 months old were not included in these analyses.

2.2.1.4 | Long-term match

Matches that lasted 24 months or longer were coded as being long-term matches. Matches that closed prior to 24 months were coded as not being long-term matches. Open matches that were less than 24 months old (i.e., they did not have the opportunity to be a long-term match) were not included in the analyses related to this outcome.

2.2.2 | Program practices

Program practices were assessed using a shortened, 31-item version of the Elements Quality Improvement Process (EQUIP) program self-assessment questionnaire (Kupersmidt, Stelter, & Rhodes, 2011). In some cases, more than one question was used to assess implementation of a practice. The items were combined into 22 practices, based upon the benchmark practices included in the EEPM (MENTOR, 2009). Mentoring staff reported on whether or not their programs were implementing each practice and their responses were coded as either implementing the practice or not.¹ A total benchmark score was calculated as the sum total number of practices implemented ranging from none to 22.

The benchmark practices were further reduced to six standards, organized according to the life cycle of a mentoring relationship. The standards include recruitment (two benchmarks), screening (eight benchmarks), training (two benchmarks), matching (two benchmarks), monitoring and support (five benchmarks), and closure (three benchmarks). Programs were coded as implementing the standard if they endorsed all of the benchmark practices that compose the EEPM-defined standard. A total standard score was calculated as the sum total number of standards implemented of the six total possible standards.

2.3 | Procedure

The BBBSA AIM archival database was transferred through a secure file transfer protocol for data analysis purposes. The database was anonymized and prepared for analysis. Each youth was represented in the analysis data set once.

¹ The EEPM has been recently updated to a fourth edition; however, data collection for the present study was designed and completed prior to the release of the fourth edition. Hence, the results reported in this paper are based upon the benchmark practices described in the third edition. All of the benchmark practices in the third edition were retained in the fourth edition.

Data cleaning procedures included checking to identify youth who had been rematched or appeared multiple times in the data set. When youth appeared in the data set more than once, only data and match characteristics were retained for the relationship with the earliest match start date. The data set included matches from school-based and community-based programs; however, because of the analytical focus on match length as an outcome, only community-based matches were retained for analysis, as community-based programs expect longer match lengths.

Executive directors (EDs) of a subsample of BBBS agencies provided consent for their agencies to participate in the study. EDs either completed the online survey themselves or suggested that someone else in their agency complete the survey. Participants included knowledgeable staff members who were responsible for managing program services at their agencies (e.g., program manager, vice president of operations).

Participants first endorsed an online informed consent form, and then responded to the 45-minute online questionnaire regarding their program's practices. After submitting their online questionnaire, programs were paid a small incentive for their participation.

3 | RESULTS

3.1 | Overview of the analyses

Six sets of analyses were conducted. First, descriptive statistics were calculated to describe the prevalence of implementation of the six standards and 22 EEPM benchmark practices reported by participating programs. Second, preliminary multilevel regression analyses were conducted to examine level 1 and level 2 characteristics of mentees and programs that might need to be included as covariates in the main analyses. Third, five standards were examined as predictors of match length in multilevel regression analyses to assess if any standard was individually predictive. Fourth, the sum total number of benchmarks and the sum total number of standards for each program were examined in separate analyses as predictors of match length using multilevel regression analyses. Fifth, match length was assessed in two survival analyses to determine whether programs that adhered to more benchmarks or more standards had longer matches than other programs. Finally, multilevel logistic regressions were employed to determine whether cumulative program practice implementation was associated with premature match closure and long-term match length.

3.2 | Prevalence of implementation of benchmark practices

The number of benchmark practices implemented by participating programs ranged from a low of 13 to a high of 21 practices of the 22 practices that were assessed. Programs reported executing an average of 17 benchmark practices. The percentage of programs that reported adhering to each benchmark practice can be seen in Table 1. Mentoring programs implemented the largest percentage of research-based practices for the screening, matching, and closure standards with 60% to 70% compliance. About a third of programs implemented all of the training benchmarks. The standards that were least frequently fully implemented were recruitment, and monitoring and support. Only one program was in full compliance with the EEPM monitoring and support standard, which was primarily due to the fact that the BBBSA standard of practice at the time of the study for monitoring matches was quarterly contact after one year of a mentoring relationship, as opposed to the EEPM benchmark of monthly contact. Therefore, due to lack of variance, the EEPM monitoring and support standard was not tested as an individual predictor in any analyses.

In terms of implementation of the individual benchmark practices, all programs implemented several of the screening benchmarks related to obtaining written documentation and contact information from mentors and families. In addition, all programs reported implementing practices involving obtaining criminal background and reference checks of mentors. All programs arranged and documented the first meetings of all matches, and all programs provided resources to postmatch mentors to help them overcome ongoing challenges to mentoring. In contrast, a very small percentage of programs reported maintaining monthly contact with matches throughout the life of the match; documenting information such as dates, lengths, and the nature of all match meetings; and employing recruitment strategies that realistically portrayed mentoring to potential mentors and mentees.

TABLE 1 Percentage of Mentoring Programs (n = 45) Reporting Implementation of Benchmark Practices

Standard/ benchmark	Brief practice description	Percentage reporting implementation ^a
Recruitment		13
B.1.1	Recruitment strategies realistically portraying mentoring to prospective mentors	42
B.1.2	Recruit youth to be mentees that match program services and communicate what mentoring is and what they can expect from a mentoring relationship	26
Screening		71
B.2.1	Mentors complete an application	100
B.2.2	Mentor agrees to a one (calendar or school) year minimum commitment	100
B.2.3	Mentor agrees to minimum frequency and duration of match meetings	75
B.2.4	Program conducts at least one face-to-face interview of prospective mentor	98
B.2.5	Program conducts reference check on prospective mentor	100
B.2.6	Program conducts a comprehensive criminal background check on prospective adult mentors	100
B.2.7	Parent or guardian completes an application and provides permission for child participation in program	100
B.2.8 and B.2.9	Parent or guardian and mentee agree to a one (calendar or school) year minimum commitment; parent or guardian and mentee agree to a minimum frequency and duration of match meetings	96
Training		38
B.3.1	Mentors receive a minimum of two hours of pre-match, in-person training	64
B.3.2	Mentor training includes a minimum of seven research-based topics	49
Matching		78
B.4.1	Mentor and mentee characteristics such as shared interests considered when making matches	78
B.4.2	Program facilitates first match meeting	100
Monitoring and support		2
B.5.1	Program contacts each mentor and mentee at a minimum frequency of twice in the first month and then, monthly thereafter	13
B.5.2	Program documents information about match meetings	22
B.5.3	Program provides mentors with resources to help handle relationship challenges	100
B.5.4	Program assesses match relationship using scientifically-tested tools	96
B.5.5	Program provides post-match training	87
Closure		67
B.6.1	Program has protocol for managing anticipated closures	98
B.6.2	Program has protocol for managing unanticipated closures	93
B.6.3	Program conducts exit interviews with match members	69

^aPercentage for each standard refers to the percentage of programs that reported implementing all of the benchmark practices that define a standard.

* $p < 0.05$.

3.3 | Preliminary match length analyses

A preliminary set of multilevel models were estimated to first calculate the intra-class correlation for match length and second, to examine the degree to which demographic characteristics of mentees and matches were associated with match length, while controlling for the nestedness of data within program. The multilevel model building procedure began with an unconditional model to determine whether agencies differed from each other in terms of match length. According to the intercept of the unconditional model, matches lasted approximately 19.2 months. However, the model also yielded a significant variance of the random intercept, suggesting that agencies differed from one another in how long their matches lasted, on average. Based on the variance components from the unconditional model, the intraclass correlation for match length was 0.03, suggesting that 3% of the total variance in match length is accounted for by differences between agencies in their average match length.

After estimating the unconditional model, level 1 covariates were added to the model one at a time and retained in the model if they were significant. A nominal four-level cohort variable was examined as the first covariate added to the model, because individuals from earlier matches, by definition, had more opportunity to be part of a longer match. The most recent cohort served as the referent group. As expected, earlier cohort groups had significantly longer matches. Thus, the cohort variables were retained as covariates in all further models.

Fixed effects for mentees' age, sex, race, and ethnicity were added consecutively to a series of models, and variables were retained if they were significantly associated with match length. Age was mean-centered (mean = 11.01 years old). Results indicated that age and sex should be retained as covariates, whereas race and ethnicity should not. Mentees who were older when they were matched experienced slightly shorter matches, and boys experienced slightly longer matches than girls. Age was included as a random slope and results indicated that the variance across agencies in the effect of age was statistically significant ($\hat{\tau}_{11} = 0.025$); therefore, there is evidence that the effect of mentee age on match length varies across agencies. Finally, a level 2 fixed effect of agency size was added to the model to determine whether bigger or smaller agencies differed in their average match lengths. Agency size was mean-centered (mean = 651.39) and was not a significant covariate. The final baseline covariate model can be seen in Table 2.

3.4 | Individual program practice standards as predictors of match length

Parameter estimates for the effect of program practices on match length are included in Table 2 in the columns following the baseline covariate model. Each Standard was tested as a predictor of match length by independently adding each one to the baseline covariate model, checking for significance of the fixed parameter estimate ($\hat{\gamma}_{01}$) and comparing the deviance (-2Log Likelihood) estimates between the models, which is χ^2 distributed, with 1 degree of freedom. In addition, preliminary exploratory models included interactions between cohort and program practices predicting match length. None of the interactions were significant and the interactions were removed from the final models.

The training standard was the only standard that was significantly associated with match length ($\hat{\gamma}_{01} = 1.44, p < .01$) and, based on comparisons of the -2LL values, the model significantly differed from the baseline covariate model, $\chi^2(1) = 7.6, p < .01$. After including covariates, adding training to the model did not account for any of the remaining level 1 variance, but it did account for approximately 10% of the remaining level 2 variance ($1 - 2.589/2.891$) between agencies.

3.5 | Sum total implementation of benchmark and standard practices as predictors of match length

Parameter estimates for the effect of sum total benchmarks and standards on match length can be seen in the right-most columns of Table 2. The sum total number of benchmarks and sum total number of standards are both mean-centered (mean of total benchmarks = 17; mean of total standards = 3). Both benchmarks ($\hat{\gamma}_{02} = 0.39$) and standards ($\hat{\gamma}_{02} = 0.54$) were significantly associated with match length and, by comparing the -2LL values of the models to that

TABLE 2 Parameter Estimates for Models Examining the Associations between Match Length and Individual Standards and Total Program Practices

	Baseline Covariate Model	Standard (\hat{y}_{01})						Total Program Practices (\hat{y}_{02})	
		Recruitment	Screening	Training	Matching	Closure	Total Benchmarks	Total Standards	
Fixed components									
Intercept	14.56***	14.48***	14.13***	14.07***	14.19***	14.34***	14.77***	14.90***	
Cohort 1 (pre-2009)	8.74***	8.74***	8.74***	8.74***	8.74***	8.74***	8.52***	8.52***	
Cohort 2 (2009)	7.15***	7.14***	7.14***	7.15***	7.14***	7.14***	6.98***	6.98***	
Cohort 3 (2010)	4.97***	4.97***	4.97***	4.97***	4.97***	4.97***	4.88***	4.88***	
Age	-0.41***	-0.41***	-0.41***	-0.41***	-0.41***	-0.41***	-0.40***	-0.41***	
Sex	1.50***	1.49***	1.50***	1.50***	1.50***	1.50***	1.45***	1.45***	
Standard		0.64	0.65	1.44**	0.54	0.36			
Total Practices							0.39*	0.54*	
Variance of random components									
$\hat{\sigma}_0$	2.891***	2.939***	2.840***	2.589***	2.918***	2.942***	2.152***	2.192***	
$\hat{\sigma}_{11}$	0.025*	0.025*	0.025*	0.025*	0.025*	0.025*	0.029*	0.028*	
$\hat{\sigma}_{01}$	-0.038	-0.025	-0.032	-0.049	-0.028	-0.038	0.019	0.00	
σ^2	139.46	139.45	139.46	139.45	139.45	139.46	139.36	139.36	
Deviance (-2LL)	234345.6	234343.6	234343.6	234338.0	234344.0	234344.5	217558.9	217559.9	

Notes: Age, Total Benchmarks, and Total Standards were mean-centered. Referent group for Cohort is Cohort 4 (2011-2012). Referent group for Sex is female. The Monitoring and Support Standard was not tested as an individual predictor, due to lack of variance. It was included in calculations for Total Benchmarks and Total Standards.

* $p < .05$. ** $p < .01$. *** $p < .001$.

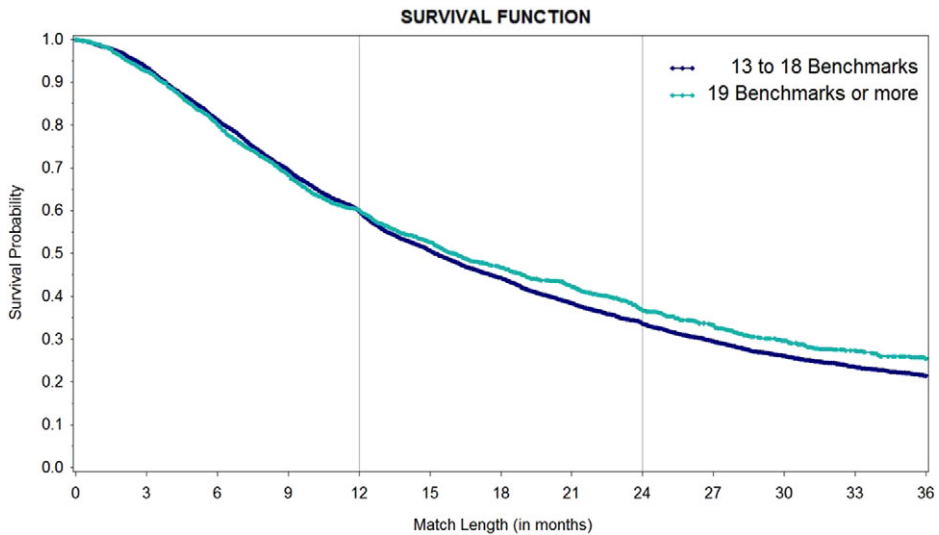


FIGURE 1 Survival analysis of match length by benchmark implementation groups

of the baseline covariate model, both total number of benchmarks, $\chi^2(1) = 16786.1$, $p < .001$, and total number of standards, $\chi^2(1) = 16785.7$, $p < .01$, were significantly different from the baseline model. As with the variance components for the training standard, total benchmarks and total standards did not account for any of the remaining level 1 variance, after the inclusion of covariates. However, adding total benchmarks to the model accounted for 26% of the remaining level 2 variance of match length, and total standards accounted for 24% of the remaining level 2 variance.

3.6 | Survival analyses of sum total benchmarks and sum total standards implemented predicting match length

Benchmarks and standards implementation were each examined as predictors of match length in separate survival analyses (Singer & Willet, 1991) to assess whether strong program adherence was associated with different relationship survival patterns. Analyses were conducted using PROC LIFETEST in the SAS program. The dichotomous variable denoting whether or not matches were open or closed at study termination was employed as a censoring variable to indicate that active matches have unknown true match lengths, but that they lasted at least as long as the recorded length of the study.

To test the effect of the sum total number of benchmarks endorsed on match length, programs were divided into two strata: those reporting high levels of implementation (endorsing 19 or more benchmarks; 20% of programs) and those reporting average to low levels of implementation (endorsing 18 or fewer benchmarks; 80% of programs). The cumulative survivor functions for benchmark implementation groups with surviving matches are presented in Figure 1. The survivor curve for the high implementation group is significantly higher than that of the low implementation group, particularly for longer lasting matches. The survival analysis indicated that the difference between Benchmark group strata was statistically significant, log-rank chi-squared (1) = 4.69, $p < .05$.

To test the effect of the sum total number of standards endorsed on match length, programs were again divided into two strata: those that were high on implementation (endorsing four or more standards; 27% of programs) and those that were average to low on implementation (endorsing three or fewer standards; 73% of programs). The survivor functions for the two standards groups are illustrated in Figure 2. As seen in the analysis of benchmark compliance, the survivor curve for the high implementation group was consistently higher than that of the low implementation group, and the difference is more pronounced for longer matches. Comparisons between the strata indicate that the difference between standards implementation groups is statistically significant, log-rank chi-squared (1) = 5.69, $p < .05$.

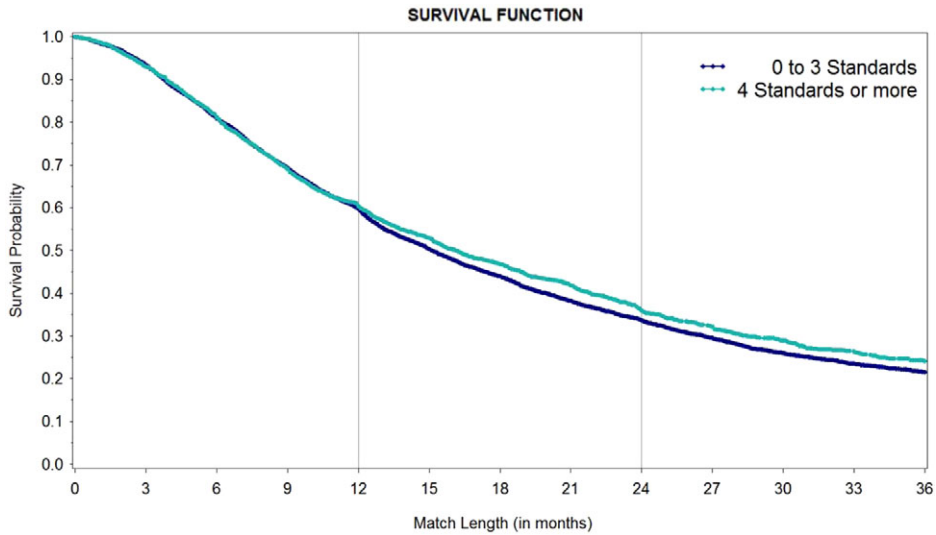


FIGURE 2 Survival analysis of match length by standards implementation groups

3.7 | Sum total implementation of benchmark and standard practices predicting premature closure and long-term match longevity

A series of multilevel logistic regressions were estimated to determine whether implementation of benchmarks and standards were associated with premature match closure (matches ending before the expected 12 month commitment) and long-term match survival (matches lasting longer than 24 months). Preliminary covariate models were built using the same procedure that was described for the match length analyses. As with the match length analyses, parameter estimates for cohort, age, and sex were significant as fixed effects, though age did not significantly add to the models when included as a random slope in the premature closure and long-term match analyses. Parameter estimates and odds ratios for baseline covariate models for the premature closure and long-term match analyses are included in Tables 3 and 4, respectively.

For premature closure, the intercept parameter estimate of -0.93 in the baseline covariate model refers to matches in the most recent cohort with female, average age mentees (11 years old), which translates to 28% probability of premature closure, $e^{(-0.93)} / (1 + e^{(-0.93)})$. According to the covariate model, probabilities of premature closure are larger for earlier cohorts and older mentees, and smaller for male mentees. Neither the total number of benchmarks nor the total number of standards implemented by participating programs significantly added to the prediction of premature closure.

For the long-term match analyses, the intercept parameter estimate -1.65 translates to a 16% probability of being in a long-term match, for 11-year-old girls in the most recent cohort. The lower probability for long-term matches in the most recent cohort is likely an artifact of the reduced sample size of the cohort. The probabilities of being in a long-term match are much higher for earlier cohorts, ranging from 34% to 37%. Boys were more likely and older mentees were less likely to be part of a long-term match. Both the total number of benchmarks and total number of standards implemented were significantly associated with being part of a long-term match.

Probabilities for 11-year-old girls in programs implementing an average number of benchmarks ranged from 17% to 38%, depending on cohort, and increase to 18% to 40%, with the addition of one more implemented benchmark. With the addition of three more implemented benchmarks, probabilities increased from 21% to 44%, with additional adjustments based on age and sex. Probabilities for 11-year-old girls in programs implementing an average number of standards ranged from 17% to 38%, depending on cohort, and increased from 19% to 41%, with the addition of one more implemented standard. With the addition of two more implemented benchmarks, probabilities increased from 21% to 45%, with additional adjustments based on the age and sex of the mentee.

TABLE 3 Parameter Estimates for Logistic Models Examining the Associations between Premature Match Closure and Program Practices

		Baseline Covariate Model		Total Program Practices (\hat{y}_6)			
				Total Benchmarks		Total Standards	
		Estimate	OR	Estimate	OR	Estimate	OR
Fixed components							
Intercept	\hat{y}_0	-0.93***		-0.97***		-0.97***	
Cohort 1 (pre-2009)	\hat{y}_1	0.20***	1.22	0.22***	1.24	0.21***	1.24
Cohort 2 (2009)	\hat{y}_2	0.18***	1.20	0.19***	1.21	0.19***	1.21
Cohort 3 (2010)	\hat{y}_3	0.10**	1.11	0.11***	1.12	0.11**	1.12
Age	\hat{y}_4	0.06***	1.06	0.06***	1.06	0.06***	1.06
Sex	\hat{y}_5	-0.16***	0.85	-0.15***	0.86	-0.15***	0.86
Total Program Practices	\hat{y}_6			-0.04	0.96	-0.03	0.98
Variance of random intercept							
	$\hat{\sigma}_0^2$	0.107*		0.098*		0.101*	

Notes. The Premature Match Closure variable was defined as matches with lengths less than 12 months old. Age, Total Benchmarks, and Total Standards were mean-centered. Referent group for Cohort was Cohort 4 (2011-2012). Referent group for Sex is female. The Monitoring and Support Standard and Benchmarks were included in calculations for Total Benchmarks and Total Standards.

* $p < .05$. ** $p < .01$. *** $p < .001$.

4 | DISCUSSION

This study was the first to examine the relationship between the comprehensive set of standards and benchmark practices for the mentoring field with measures of match longevity. Overall, the results revealed that the sum total number of benchmarks and standards implemented by mentoring programs was significantly associated with match length, particularly for matches surviving for 2 years or longer. The only individual standard that independently predicted match length was training. These findings suggest that greater adherence to and fidelity of implementation with the EEPM (MENTOR, 2009) can enhance program effectiveness.

The level of adherence to practice guidelines matters. As the number of benchmark practices increase, relationship longevity and survival increase. These findings on the importance of adhering to core evidence-based practices are consistent with those reported in other practice literatures. For example, in the juvenile justice literature, programs following the best practice guidelines when implementing generic juvenile justice interventions are as effective in achieving reductions in recidivism as programs that faithfully implement "certified" model interventions (Howell & Lipsey, 2012). Given findings that relationship quality and longevity are the active ingredients in effective mentoring programs (Bayer, Grossman, & DuBois, 2015), the approach of mentoring programs implementing generic evidence-based practices may provide a complementary approach to the use of model interventions and allow the field of mentoring to take effective programming to scale.

The diversity in program practice implementation reported by mentoring programs in this study is particularly notable, given that all of the participating programs are affiliate members of BBBSA and have many common goals, core tools, staff professional development and certification training guidelines, and service delivery requirements. Despite a shared vision, mission, and goals, these data demonstrate that there was still meaningful variation across BBBS programs in their everyday implementation of program practices. For example, whereas 100% of program leaders reported implementing practices designed to protect the physical safety of mentees (e.g., conducting criminal background and reference checks of volunteer mentors), only a small percentage reported employing recruitment strategies that realistically portray mentoring to prematch mentees. When respondents were asked whether incoming mentees understood what it meant to be mentored, the majority of them reported that they did not.

TABLE 4 Parameter Estimates for Logistic Models Examining the Associations between Long-Term Matches and Program Practices

		Baseline Covariate Model		Total Program Practice (\hat{y}_6)			
				Total Benchmarks		Total Standards	
		Estimate	OR	Estimate	OR	Estimate	OR
Fixed components							
Intercept	\hat{y}_0	-1.65***		-1.62***		-1.59***	
Cohort 1 (pre-2009)	\hat{y}_1	1.01***	2.74	1.01***	2.74	1.01***	2.74
Cohort 2 (2009)	\hat{y}_2	1.06***	2.90	1.06***	2.89	1.06***	2.89
Cohort 3 (2010)	\hat{y}_3	1.12***	3.08	1.12***	3.07	1.12***	3.07
Age	\hat{y}_4	-0.06***	0.94	-0.06***	0.94	-0.06***	0.94
Sex	\hat{y}_5	0.28***	1.32	0.27***	1.31	0.27***	1.31
Total Program Practices	\hat{y}_6			0.09***	1.10	0.13**	1.13
Variance of random intercept							
	$\hat{\sigma}_0^2$	0.126*		0.083*		0.086*	

Notes. The Long-term Matches variable was defined as matches that lasted 24 months or longer. Age, Total Benchmarks, and Total Standards were mean-centered. Referent group for Cohort was Cohort 4 (2011-2012). Referent group for Sex is female. The Monitoring and Support Standard and Benchmarks were included in calculations for Total Benchmarks and Total Standards.

* $p < .05$. ** $p < .01$. *** $p < .001$.

There was one service delivery model that BBBS programs used during the duration of the study and it involved a set of program practices that were designed to support the creation of positive, community-based matches that would last for a minimum of one calendar year. An example of a screening practice designed to achieve that goal was to ask volunteers to make a 1-year commitment. The fact that the logistic regression analysis predicting premature match closure from EEPM program practices was not significant suggests that the base BBBSA service delivery model was effective in achieving its goal of sustaining matches to last through the 12-month minimum commitment. However, as relationship longevity grew, so did the differences between agencies implementing a smaller versus larger number of benchmark practices. Higher implementation agencies were more effective in developing and sustaining longer term matches, with supporting evidence stemming from the survival analysis of long-term matches.

These data suggest that, for purposes of supporting and sustaining matches for longer periods of time, greater adherence to evidence-based practices is needed. In fact, as part of a national franchise model, a new BBBSA service delivery model (SDM) and new standards of practice for one-to-one mentoring (SOP; BBBSA, 2014) were released and implemented in January 2014, after this study had been initiated. The new SDM was updated, improved, and deployed to meet both the SOP for one-to-one mentoring and to align with the EEPM (MENTOR, 2009). Furthermore, the new SDM was designed to provide more evidence-based, predictable, and consistent service delivery across the BBBS footprint; to increase match safety, longevity, and quality; and to increase the likelihood of positive youth outcomes.

The results of this study also point to the primacy of the training standard for achieving longer matches. Training was the only EEPM standard that was independently associated with match length. Notably, programs varied on requiring prospective mentors to participate in training experiences that were inclusive of all of the topics required in the EEPM. Recent research highlights the importance of mentor training in preparing volunteers for being ready to engage in their mentoring relationships. For example, mentors who completed a web-based training program that included all topics required in the EEPM had greater knowledge, better understanding of the roles they should and should not play, and felt more prepared and ready to be a mentor than mentors who only received training-as-usual by their mentoring programs (Kupersmidt, Stelter, Rhodes, & Stump, in press).

In addition, other studies report that initial mentor training and confidence significantly predict mentor satisfaction with (Martin & Sifers, 2012) as well as commitment to (McQuillin, Straight, & Saeki, 2015) their mentoring relationships. Thus, adequate mentor preparation may be a key tool for promoting both mentor satisfaction and longevity in matches.

4.1 | Implications for mentoring program policies and practices

Practice guidelines have become popular and even mandated, across institutions, with this growth predicated on the notion that standardization of evidence-based policies and practices will improve the quality of services offered and, in turn, individual outcomes (Petrosino, Boruch, Soydan, Duggan, & Sanchez-Meca, 2001). However, despite efforts to increase adherence to practice guidelines, compliance with protocols is often incomplete or low (e.g., special education: Burns & Ysseldyke, 2009; nursing: Gotham, 2006). The broader implementation science literature can provide the mentoring field with guidance regarding strategies for promoting the universal adoption of the benchmark practices in the EEPM by mentoring programs. Common challenges to implementation of evidence-based practices that have been reported across literatures include factors such as lack of buy-in, inadequate staff training, high staff turnover, insufficient resources, inadequate infrastructure, lack of organizational support, and unclear or unrealistic expectations (Ayers & Griffith, 2007; Corrigan, Steiner, McCracken, Blaser, & Barr, 2001; Fixsen, 2012).

Despite these challenges, additional study and testing of implementation drivers such as management support (Wade & Neuman, 2007), financial resource availability (Dappen & Isernhagen, 2005), staff training and supervision (Gottfredson & Gottfredson, 2002; Keller, 2007), and organizational climate and attitudes that support the implementation of new program practices (Klein, Conn, & Sorra, 2001) are needed to support the adoption of the EEPM practices by individual mentoring programs and practitioners (Fixsen, 2012).

Furthermore, adoption of the Fixsen, Blase, Naoom, and Wallace (2009) model regarding the stages of implementation can further support the evolution of a mentoring program from exploration and adoption of the EEPM benchmarks to having these practices become an accepted, integrated, and sustained part of its service delivery system. Mentoring programs will benefit from training and technical assistance to assess, change, and update their program practices as well as from an iterative or cyclical process involving assessment, feedback, evaluation, and training from quality improvement evaluators and providers (Wandersman et al., 2006).

Mentoring programs and mentors should be prepared to identify and serve all mentees, especially high-risk mentees. There are several ways that programs can be prepared and competent to do so. First, prematch mentor training, as outlined in the training and closure standards in the EEPM (MENTOR, 2015), can help decrease unnecessary and early termination from occurring that stems from factors such as unrealistic expectations or poor communication skills. Mentoring a youth with multiple vulnerabilities can result in a mentor losing interest or motivation to sustain the relationship or feeling overwhelmed or ineffective. Proper training can help mentors prepare for the inevitable challenges that face all relationships, especially those with youth in need (MENTOR, 2015). Also, consistent with research and theory on the relationship between organizational support and volunteer retention (Omoto & Snyder, 1995), ongoing postmatch training can also contribute to more effective, more enduring, and higher quality mentoring relationships (DuBois et al., 2002; Herrera et al., 2007).

In addition to the benchmark practices required to meet the EEPM standards, the EEPM also suggests several additional practices that are hypothesized to enhance program effectiveness even further. Although these practices have not yet been empirically evaluated, there is a strong logical or theoretical basis for them. For example, although an enhancement in the EEPM, mentee and parent/guardian training is hypothesized as being important for enhancing understanding and commitment to the mentoring relationship, based upon findings from a series of exploratory interviews with volunteers and youth in successful and terminated relationships (Spencer, 2006). Mentees and families who are not trained on the realities of mentoring may experience disappointment and, in turn, may sabotage or even prematurely terminate the mentoring relationship. Prematch training of mentees and their parents or guardians with the skills, knowledge, and cognitions needed to develop strong relationships with mentors and mentoring program staff may strongly contribute to relationship success (MENTOR, 2015).

Another example of an enhanced practice is to reduce the staff-to-match ratio for high-risk matches, so that staff can provide more frequent and more intense match support. To support high-risk matches, staff will need more evidence-based and ongoing professional development and supervision. In fact, research suggests that match monitoring and mentor support are associated with stronger and more enduring mentoring relationships (DuBois et al., 2002). In further support of this enhanced practice, recent findings suggest that mentors who had more organizational contacts in terms of more match support staff members who communicated with the mentor and more face-to-face (as opposed to telephone and e-mail) contacts with match support staff members at their mentoring program were less likely to prematurely terminate their mentoring relationship (Stukey, 2016). Thus, once a mentoring program is adhering to the EEPM benchmark practices, further quality improvement efforts might then focus on adoption of the enhancement practices.

4.2 | Limitations and future directions

One of the limitations of the current study was that it used a correlational rather than experimental design placing limitations on conclusions about program practices being causally related to match outcomes. A second limitation is that although match longevity was examined in the present study, match quality and youth outcomes were not. Given the relationship between match quality and youth outcomes (Chan et al., 2013), studying the relation between program practices and mentor and mentee reports of match quality is a useful direction for future research. Unfortunately, match quality and youth outcome data could not be examined within the current dataset because these data were available only for mentees who were still participating in mentoring relationships at 12 months. Youth outcome data were not collected by BBBS for matches that closed prematurely. Thus, the results of analyses of youth outcomes would be biased, despite the large sample size. Other standards besides the training standard may prove to be independently related to other types of match and youth outcomes.

A third limitation is that adherence to evidence-based program practices was examined only for programs that were part of the BBBSA national model. Examining the impact of program practice implementation across a wider range of programs that use other conceptual frameworks and practice models will help to understand both variability in the field and the generalizability of the findings. Furthermore, the sample represents a convenience sample; therefore, the prevalence rates for adherence to benchmark practices may not be representative of all BBBS agencies and needs to be replicated. Nonetheless, the descriptive information included in this paper provides valuable information about natural variation in conformity to standards of practice that has been lacking in the field.

Fourth, the third edition of the EEPM (2009) is now somewhat dated with the recent release of the fourth edition of the EEPM (2015) and the new BBBSA standards of practice (2014). The fourth edition contains an additional 25 benchmarks for a total of 48 safety- and evidence-based benchmark practices. The impact on outcomes for this expanded set of practices needs to be evaluated in future research, in terms of examining the empirical basis for both each individual benchmark practice and the set of practices as a whole. Fifth, future research is needed on the quality of practice implementation (i.e., not just presence or absence of a practice) as well as actual compliance with the implementation of each benchmark practice, as it is defined, and its relationship to match and youth outcomes. Finally, measures development using other informants (e.g., using external, objective reviewers with assessment of inter-rater reliability) to assess implementation of benchmark practices will further advance the field.

4.3 | Conclusion

In conclusion, this study has made significant contributions to our understanding of the effect of program practices on match length. Many of the benchmark practices can provide better preparation and support to mentors, as well as mitigating against the frustrations or self-doubts that may arise over the natural course of building a new, helping, volunteer relationship with a child or adolescent than is typically found in the practice field. Adherence to evidence-based benchmark practices has the potential of creating more satisfying, longer lasting, and effective mentoring relationships. Universal adoption of the EEPM compendium of safety and evidence-based practices will

represent a paradigm shift from simply relying on practice wisdom to incorporating science into infrastructure, operations, and management of mentoring programs with the potential of profound impact on the development of millions of youth.

REFERENCES

- Ayres, C. G., & Griffith, H. M. (2007). Perceived barriers to and facilitators of the implementation of priority clinical preventive services guidelines. *American Journal of Managed Care*, 13(3), 150–156.
- Bayer, A., Grossman, J. B., & DuBois, D. L. (2015). Using volunteer mentors to improve the academic outcomes of underserved students: the role of relationships. *Journal of Community Psychology*, 43(4), 408–429.
- Bernstein, L., Rappaport, C. D., Olsho, L., Hunt, D., & Levin, M. (2009). Impact evaluation of the US Department of Education's student mentoring program. Final Report. NCEE 2009-4047. *National Center for Education Evaluation and Regional Assistance*.
- Big Brothers Big Sisters of America. (2014). *Service delivery model (SDM) and standards of practice for one-to-one mentoring (SOP)*. Dallas, TX: Big Brothers Big Sisters of America.
- Burns, M. K., & Ysseldyke, J. E. (2009). Reported prevalence of evidence-based instructional practices in special education. *Journal of Special Education*, 43, 3–11. doi:10.1177/0022466908315563
- Chan, C. S., Rhodes, J. E., Howard, W. J., Lowe, S. R., Schwartz, S. E., & Herrera, C. (2013). Pathways of influence in school-based mentoring: The mediating role of parent and teacher relationships. *Journal of School Psychology*, 51(1), 129–142.
- Corrigan, P. W., Steiner, L., McCracken, S. G., Blaser, B., & Barr, M. (2001). Strategies for disseminating evidence-based practices to staff who treat people with serious mental illness. *Psychiatric Services*, 52(12), 1598–1606.
- 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*, 31(2), 157–197.
- 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, 57–91.
- Fixsen, A. A. M. (2012). *Implementer perspectives: The implementation of a school-based mentoring program* (Doctoral dissertation). Portland State University, Portland, OR.
- Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F. (2009). Core implementation components. *Research on Social Work Practice*, 19(5), 531–540.
- Gotham, H. J. (2006). Advancing the implementation of evidence-based practices into clinical practice: How do we get there from here? *Professional Psychology: Research and Practice*, 37(6), 606.
- Grossman, J. B., Chan, S. C., Schwartz, S. E., & Rhodes, J. E. (2011). 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 programs. *American Journal of Community Psychology*, 3, 199–219.
- Herrera, C., Grossman, J. B., Kauh, T. J., Feldman, A. F., McMaken, J., & Jucovy, L. Z. (2007). *Making a difference in schools: The Big Brothers Big Sisters school-based mentoring impact study*. Philadelphia: Public/Private Ventures.
- Howell, J. C., & Lipsey, M. W. (2012). Research based guidelines for juvenile justice programs. *Justice Research and Policy*, 14(1), 17–34.
- Kupersmidt, J. B., Stelter, R. L., & Rhodes, J. E. (2011). *Elements quality improvement process (EQUIP): Web-based, program self-assessment questionnaire*. Durham, NC: Innovation Research and Training.
- Kupersmidt, J. B., Stelter, R. L., Rhodes, J. E., & Stump, K. N. (in press). Enhancing mentor efficacy and preparedness through Web-based, pre-match training. Manuscript submitted for publication. *Journal of Nonprofit Education and Leadership*.
- Kupersmidt, J. B., Stump, K. N., Stelter, R. L., & Rhodes, J. E. (in press). Predictors of premature match closure in youth mentoring relationships. Manuscript accepted for publication, *American Journal of Community Psychology*.
- Martin, S. M., & Sifers, S. K. (2012). An evaluation of factors leading to mentor satisfaction with the mentoring relationship. *Children and Youth Services Review*, 34(5), 940–945.
- McQuillin, S. D., Straight, G. G., & Saeki, E. (2015). Program support and value of training in mentors' satisfaction and anticipated continuation of school-based mentoring relationships. *Mentoring & Tutoring: Partnership in Learning*, 1–16.
- MENTOR. (2009). *Elements of effective practice for mentoring* (3rd ed.). Alexandria, VA: author.
- MENTOR. (2015). *Elements of effective practice for mentoring* (4th ed.). Boston, MA. MENTOR. (2009). *Elements of effective practice for mentoring* (3rd ed.). Alexandria, VA: author.

- Omoto, A. M., & Snyder, M. (1995). Sustained helping without obligation: motivation, longevity of service, and perceived attitude change among AIDS volunteers. *Journal of Personality and Social Psychology*, *68*(4), 671.
- Petrosino, A., Boruch, R. F., Soydan, H., Duggan, L., & Sanchez-Meca, J. (2001). Meeting the challenges of evidence-based policy: The Campbell Collaboration. *The ANNALS of the American Academy of Political and Social Science*, *578*(1), 14–34.
- Rhodes, J. E., Schwartz, S. E., Willis, M. M., & Wu, M. B. (2014). Validating a mentoring relationship quality scale: Does match strength predict match length? *Youth & Society*. doi:10.1177/0044118X14531604
- Singer, J. D., & Willett, J. B. (1991). Modeling the days of our lives: Using survival analysis when designing and analyzing longitudinal studies of duration and the timing of events. *Psychological Bulletin*, *110*(2), 268.
- Spencer, R. (2006). Understanding the mentoring process between adolescents and adults. *Youth & Society*, *37*, 287–315.
- Stukey, J. K. (2016). Predictive factors of organizational support communication in volunteer mentor retention (Doctoral dissertation). Walden University, Minneapolis, MN.
- Wandersman, A., Clary, E. G., Forbush, J., Weinberger, S. G., Coyne, S. M., & Duffy, J. L. (2006). Community organizing and advocacy: Increasing the quality and quantity of mentoring programs. *Journal of Community Psychology*, *34*(6), 781–799.
- 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*. Philadelphia: Big Brothers Big Sisters of America.

How to cite this article: Kupersmidt J. B., Stump K. N., Stelter R. L., Rhodes J. E. (2016). Mentoring program practices as predictors of match longevity. *J. Community Psychol.*, *XX*, XX–XX. DOI: 10.1002/jcop.21883