Effectiveness of mentoring programs for youth: A meta-analytic review.

American Journal of Community Psychology April , 2002

David L. DuBois (1)

INTRODUCTION

During the past decade mentoring programs for youth have become increasingly popular and widespread. Big Brothers/Big Sisters of America (BB/BSA), the most prominent of these programs, now includes over 500 agencies nationwide. The National Mentoring Partnership and numerous other organizations also have contributed to significant growth in mentoring initiatives at local, state, and national levels (Johnson & Sullivan, 1995). Currently, the National Mentoring Database lists more than 1,700 organizations that support mentoring activities (Save the Children, 1999).

Interest in mentoring programs has been fueled in significant part by the importance that positive relationships with extrafamilial adults have been indicated to have in promoting resiliency among youth from at-risk backgrounds (Rhodes, 1994). It should not be assumed, however, that the essential features of these types of naturally occurring relationships can reliably be reproduced by programs that seek to provide youth with adult mentors through necessarily more artificial mechanisms (Hamilton & Hamilton, 1992). Studies evaluating the benefits of mentoring programs for youth have begun to appear only recently in the literature. Prior reviews (Darling, Hamilton, & Niego, 1994; Flaxman, Ascher, & Harrington, 1988; Johnson & Sullivan, 1995; Rhodes, 1994), therefore, have been limited by a lack of available data upon which to base conclusions. Furthermore, because of multidisciplinary and applied interest in mentoring, reports have appeared in diverse literatures and a significant proportion have been published privately by foundations and other organizations.

The present research utilizes meta-analysis to review and synthesize the existing empirical literature on youth mentoring programs (Cooper, 1998; Durlak & Lipsey, 1991). Meta-analysis offers several advantages over the narrative approach that has been employed in prior reviews. These include (a) explicit operationalization of literature search procedures to help reduce omissions or bias in the investigations that are identified for review; (b) an objective and quantifiable basis for assessing the overall magnitude of program effects on youth; and (c) the ability to test for significant differences in findings across investigations along any dimension of interest, thus facilitating identification of factors that may have important implications for program effectiveness. This latter concern seems particularly germane to the study of youth mentoring programs given the considerable diversity that has characterized intervention efforts in this area (Rhodes, 1994). Factors meriting consideration as sources of influ ence on the results of mentoring program evaluations include (a) features of program design and implementation; (b) characteristics of participating youth; (c) qualities of the mentor-mentee relationships that are formed; and (d) issues relating to the assessment of youth outcomes.

Program Design and Implementation

From a program design standpoint, many programs (e.g., BB/BSA) have focused solely on providing mentoring relationships to youth. In other instances, mentoring has been implemented as one of several distinct components of a multifaceted intervention program. Enhanced benefits generally have been expected to result when mentoring is linked to other supportive services (Flaxman et al., 1988). Nevertheless, there also may be certain advantages to program specialization in mentoring. With regard to this latter possibility, BB/BSA has been widely discussed as a model of "best practices" for youth mentoring (e.g., Tierney, Grossman, & Resch, 1995). The effectiveness of this program relative to non-BB/BSA programs is thus of particular interest.

Mentoring programs also have differed in their basic goals and philosophy. Thus, whereas some programs have pursued the general goal of promoting positive youth development, others have adopted more focused or instrumental goals relating to areas such as education or employment (Saito & Blyth, 1992). The relative merits of these contrasting program orientations has attracted a considerable amount of

Assessment of Outcomes

Type of outcome assessed was not a significant moderator of effect size (see Table II). Under the assumption of fixed effects, the 95% confidence intervals associated with effect size estimates were consistent with a positive effect of mentoring programs on all five types of outcomes examined (i.e., emotional/psychological, problem/high-risk behavior, social competence, academic/educational, and career/employment), although only to a marginal extent for emotional/psychological adjustment. Under the assumption of random effects, this was the case for three types of outcomes (i.e., problem/high-risk behavior, academic/educational, and career/employment), the exceptions being measures of social competence and emotional/psychological adjustment.

Similarly, neither data source nor timing of assessment were found to be significant moderators of effect size. Under the assumption of fixed effects, confidence intervals for effect size estimates were consistent with favorable effects of mentoring for all data sources (i.e., youth, parent, teacher, and administrative records) and for assessments occurring during programs, at immediate posttest, and at follow-up. By comparison, under the assumption of random effects this was the case only when youth constituted the data source and when assessments took place either during the program or at an immediate posttest. Length of follow-up assessment was not a significant moderator in either type of analysis, although the small number of samples involved precluded inferences of positive effects of mentoring within specific ranges of this variable (i.e., less than or equal to 1 year vs. greater than 1 year).

Analyses Controlling for At-Risk Status and Type of Outcome

As indicated previously, at-risk status was a significant moderator of effect size under the assumptions of both fixed- and random-effect analysis. It therefore was important to consider the extent to which this variable exhibited associations with other moderator variables investigated and whether or not controlling for these would have any substantial implications for primary study results. Despite the evidence to suggest that total number of theory-based or empirically based best practice indicators might vary significantly across at-risk status category, this was not found to be the case (ps > .10). Selected other variables, however, including some of the indicators that comprised each of these indices, did exhibit significant covariation with at-risk status. Illustratively, a significant association was evident between at-risk status of the sample and whether or not mentors had a background in a helping role or profession, [chi square](3) - 11.71, p < .01, with samples of youth in the individual risk sta tus category accounting for a disproportionately large proportion of the instances in which mentors with such backgrounds were used in programs (i.e., 9 of the 12 independent samples involved).

To investigate the influence of their associations with at-risk status, all remaining variables shown in Table II were reevaluated as possible moderators of effect size with statistical control for this characteristic (i.e., residualizing all effect size estimates on at-risk status of the associated sample, using three dummy variables to represent the four possible categories of risk status). Introducing this additional control produced few noteworthy changes in results. Specifically, with only two exceptions, all variables that had previously reached or approached significance as moderators in primary analyses under the assumptions of either a fixed- or random-effects model continued to do so in these supplementary analyses. The exceptions were that monitoring of implementation no longer approached significance as a moderator under the assumption of random effects, nor did low socioeconomic status under the assumption of fixed effects.

Further analyses investigated the extent to which results of moderator analyses were robust to possible confounding of the different variables involved with type of outcome measure utilized in evaluations. As noted previously, the overall degree of variation in average effect size estimates across category of outcome measure was not statistically significant. It still nevertheless was important to examine whether the variation that was evident represented a source of influence on the findings of other analyses (cf. Durlak & Wells, 1997). Results of analyses that included control for type of outcome assessed revealed only a few substantive changes from those reported in primary analyses. Specifically, as was the case when controlling at-risk status, low socioeconomic status no longer approached significance as a moderator under the assumption of fixed effects. Program goal and screening of prospective mentors also now approached significance as moderators in fixed-effect analyses (ps < .10). These latter findi ngs involved the same trends that are evident in Table II toward larger estimates of effect size for those mentoring programs that emphasized instrumental goals for youth and those that indicated use of procedures for screening prospective mentors.

Intervention Group Comparisons on Relationship Quality

The final set of analyses investigated effect sizes for comparisons that were made within the intervention group on the basis of relationship factors. The information needed to calculate this type of effect size was available for nine independent samples, each of which appeared in

a different study. The relationship factors assessed in these reports included longevity (Royse, 1998), frequency and amount of contact (Howitt, Moore, & Gaulier, 1998), and whether or not a mentor was actually received within the context of the multicomponent Career Beginnings program (Cave & Quint, 1990); in the remaining studies, broader indices or categories of relationship quality were derived from sources that included mentor visit reports (Dicken, Bryson, & Kass, 1977), nominations from teachers (Huisman, 1992) or program staff (LoSciuto, Rajala, Townsend, & Taylor, 1996), and youth ratings of their experiences with mentors (Johnson, 1997; Slicker & Palmer, 1993; Stanwyck & Anson, 1989). Effect sizes were calculated for all r elevant comparisons and coded such that positive values indicated more favorable outcomes for youth experiencing greater intensity or quality of mentoring. When findings were reported as an association between a continuous relationship measure and program outcome (e.g., Pearson r), the finding reported was converted to a d-index effect size, using the appropriate formula (Cooper, 1998).

Across the nine independent samples, a total of 35 effect sizes were able to be calculated for comparisons within the intervention group on the basis of relationship factors. Following procedures described previously, distributions of effect size and sample size were inspected for outliers, with one sample size that qualified as an outlier Winsorized to a less extreme value (i.e., 300). The resulting average unweighted d-index for the 35 effect size estimates was d = .22. Using the nine independent samples involved as the unit of analysis, the average unweighted d-index was .33. When effect sizes were weighted by the inverse of their variance and a fixed-effects model was assumed, the average effect size for the nine independent samples was d = .29. Thus, on average, among youth participating in mentoring programs, those for whom relationships of greater intensity or quality were evident scored between one quarter and one third of a standard deviation higher in a favorable direction on outcome measures. The 9 5% confidence interval for this weighted d-index encompassed a lower value of d = .16 and an upper value of d = .42. Under the assumption of random effects, the average effect size for the nine independent samples was d = .30 with a 95% confidence interval extending from d = .15 to d = .45.

DISCUSSION

Findings of this investigation provide support for the effectiveness of youth mentoring programs. Results of a fixed-effects model analysis indicate an overall or average positive effect for those specific mentoring programs that have been the subject of formal evaluation (i.e., those included in the present review); a random-effects model analysis, furthermore, suggests that benefits of mentoring may generalize to a broader range of approaches to implementing this type of intervention. In accordance with the latter finding, moderator analyses revealed little evidence that the potential for programs to yield desirable outcomes is dependent on such considerations as whether or not mentoring takes place alone or in conjunction with other services, whether it is provided in accordance with the most widely implemented model (i.e., BB/BSA), or whether programs reflect relatively general (i.e., psychosocial) as opposed to more focused (i.e., instrumental) goals. Favorable effects of mentoring programs are similarly apparent across youth varying in demographic and background characteristics such as age, gender, race/ethnicity, and family structure and across differing types of outcomes that have been assessed using multiple sources of data. Although included in only a minority of studies, follow-up assessments that have been conducted also offer at least a limited basis for inferring benefits of mentoring that extend beyond the end of program participation. Cumulatively, based on available findings, it thus seems that youth mentoring programs do indeed have significant capacity to reproduce through more formal mechanisms the types of benefits that have been indicated to accrue from so-called natural mentoring relationships between youth and adults (for reviews, see Rhodes, 1994; Werner, 1995).

Results further indicate, however, that it may be most appropriate to expect the typical youth participating in a mentoring program to receive benefits that are quite modest in terms of absolute magnitude. The average estimated effect sizes of.14 and .18 obtained under the assumptions of fixed and random effects, respectively, are consistent with only a small effect for mentoring programs (Cohen, 1988; Lipsey, 1990). This degree of impact, moreover, falls substantially short of larger mean effect sizes reported previously for psychological, educational, and behavioral treatments generally (Lipsey & Wilson, 1993) and for mental health prevention programs directed at children and adolescents specifically (Durlak & Wells, 1997,1998). This aspect of findings is seemingly inconsistent with the widespread and largely unquestioned support that mentoring initiatives have enjoyed in recent years. Nevertheless, strong cautionary views have been offered previously in the youth mentoring literature (Freedman, 1992; Hamil ton & Hamilton, 1992; Rhodes, 1994). It has been pointed out in particular that numerous programmatic and other variables may be critical to attend to for the potential benefits of youth mentoring programs to be fully realized. The need for greater consideration of specific factors influencing effectiveness is underscored by the substantial overall heterogeneity in estimates of effect size observed in the present review and the numerous systematic sources of this variation that were able to be delineated in moderator analyses.

Moderators of Program Effectiveness

The theory-based and empirically based indices of best practices for mentoring programs are particularly noteworthy among the significant moderators of effect size identified. No single feature or characteristic of programs was indicated to be responsible for the positive trends in outcomes that were associated with greater degrees of utilization of either set of best practices. Several of the practices comprising the theory-based index did, however, emerge as significant individual moderators of effect size (and, hence, by definition also were included in the empirically based index), thus highlighting specific strategies that may be especially important for achieving desired results. These latter program features include ongoing training for mentors, structured activities for mentors and youth as well as expectations for frequency of contact, mechanisms for support and involvement of parents, and monitoring of overall program implementation. In multivariate analyses, these practices were further revealed to be represented consistently among the strongest predictors of greater reported positive effects for mentoring programs. The constellation of program characteristics involved reflects an emphasis on providing adequate support and structure for mentoring relationships throughout the formative strategies of their development (Hamilton & Hamilton, 1992). It is noteworthy therefore that efforts directed toward this goal apparently have been relatively neglected in youth mentoring programs to date in lieu of a greater focus on preparatory procedures such as screening, initial training and orientation, and matching of youth and mentors. Illustratively, whereas initial training or orientation has been provided to mentors on a fairly routine basis (71% of studies in the present review), efforts to provide ongoing training once relationships have begun have been much less common (23% of studies). Factors such as increased cost and reluctance to make excessive demands on volunteer mentors represent potentially formidab le obstacles to providing a more sustained infrastructure in programs (Freedman, 1992). Nevertheless, in view of available findings, it seems clear that at a minimum there is a need for decision-making in this area to incorporate careful consideration of possible implications for program outcomes.

A similarly strong linkage with beneficial outcomes is evident for the intensity and quality of relationships established between mentors and youth in programs. Specifically, among several studies in which comparisons have been made on the basis of relevant criteria within the intervention group, a substantial difference on criterion measures is apparent favoring those youth identified as having relatively strong relationships with their mentors. Many of the relationship characteristics reportedly utilized in deriving such comparisons have been found previously to be predictive of greater perceived benefits of mentoring as evaluated subjectively by mentors and youth (DuBois & Neville, 1997; Freedman, 1988; Parra et al., 1998). It appears based on this research that multiple features of relationships, such as frequency of contact, emotional closeness, and longevity, each may make important and distinctive contributions to positive youth outcomes. Unfortunately, it was not feasible to investigate this possibili ty in the present review because of the rarity with which measures of specific relationship characteristics have been included in controlled evaluations of mentoring programs. A related methodological consideration with respect to the relatively less differentiated appraisals of relationship quality that have been incorporated into existing evaluation studies is the potential for such judgements to be contaminated by knowledge of which youth mentees are prospering most in programs, thus confounding assessments of relationship factors and outcomes.

A further noteworthy result is the support found for the prevailing view that mentoring programs offer the greatest potential benefits to youth who can be considered to be at-risk (Freedman, 1992; Hamilton & Hamilton, 1992). It will be recalled in this regard that the largest estimates of effect size are evident for programs directed toward youth experiencing conditions of environmental risk or disadvantage, either alone or in combination with factors constituting individual level risk. A similar trend is apparent when considering low family socioeconomic status as a specific indicator of environmental disadvantage. Within the context of frameworks for classifying prevention efforts (Cowen, 1985; Institute of Medicine, 1994), these findings are consistent with greater effectiveness for mentoring programs characterized by a situation-focused or selective orientation. Interventions of this type focus on individuals who can be considered vulnerable by virtue of their present life circumstances, but who are not y et demonstrating significant dysfunction. Youth experiencing situations of environmental risk may be especially suitable candidates for mentoring as a preventive intervention because of a lack of positive adult support figures or role models in their daily lives (Rhodes, 1994). With respect to this possibility, available findings do not indicate reliably greater effects of mentoring for youth from single-parent households. Enhanced benefits of mentoring have been apparent in the context of low levels of perceived family support (Johnson, 1997), however, thus suggesting a need for more refined measures of risk associated with the existing support networks of youth to be included in future research. Exposure of youth to aspects of environmental adversity not assessed in evaluations could have additional significance as a factor contributing to the positive effect of mentoring that was evident to a limited degree even among those studies for which it was not possible to infer experience of any conditions of risk on the basis of the information made available.

By contrast, evidence of an overall favorable effect of mentoring is notably lacking under circumstances in which participating youth have been identified as being at risk solely on the basis of individual-level characteristics (e.g., academic failure). Mentoring is an inherently interpersonal endeavor. As a result, it may be especially susceptible to obstacles and difficulties that can arise when youth targeted for intervention are already demonstrating significant personal problems (Freedman, 1992). Many of these youth are likely to be in need of relatively extensive amounts of specialized assistance, for example, a situation that is not necessarily well-suited to the primarily volunteer

and nonprofessional status of most mentors. Considerations of this nature suggest a need for training and other appropriate forms of program support when attempting to provide effective mentoring to youth who are exhibiting individual-level risk. In accordance with this view, a more refined analysis revealed that such youth apparently can benefit significantly from participation in mentoring programs that adhere to a majority of recommended practices. Of further note are the substantial positive effects of mentoring reported for programs in which youth targeted for participation could be regarded as at-risk from both an individual and environmental perspective. Because of the relatively small number of evaluations involved, this finding merits cautious interpretation. It may be that environmental as opposed to individual risk simply has greater salience as a determining factor in likely responsiveness to mentoring. It is also possible, however, that circumstances of contextual adversity tend to reduce the likelihood of certain obstacles interfering with efforts to mentor youth who are demonstrating individual-level risk. In the presence of indications of environmental risk, for example, mentors may be less prone to accept negative labels assigned to such youth or inappropriately attribute problems they exhibit solely to persona l deficits or limitations (e.g., lack of motivation).

Applied Implications

From an applied perspective, findings offer support for continued implementation and dissemination of mentoring programs for youth. The strongest empirical basis exists for utilizing mentoring as a preventive intervention with youth whose backgrounds include significant conditions of environmental risk and disadvantage. To facilitate attainment of desired outcomes, however, results indicate a need for programs to adhere closely to recommended guidelines for effective practice (e.g., National Mentoring Working Group, 1991). Given the modest size of the effects that thus far have been able to be established for mentoring, there clearly is a rationale for innovation and experimentation with enhancements to program design. One possibility suggested by the present findings is the recruitment of mentors whose backgrounds include prior experience and success in helping roles. Older adults, for example, although underrepresented currently in programs, often may be able to bring to the mentoring role valuable skills r elating to child-rearing and other areas of life experience (Freedman, 1988; LoSciuto et al., 1996). Relative to these needs for both innovation and adherence to basic guidelines for implementation, concerns such as the most appropriate setting or goals for mentoring activities seem best to regard as being of secondary importance. Indeed, to the extent that more fundamental considerations are neglected in the development and operation of programs, there may be substantial opportunity for mentoring to have unintended negative effects on youth (Rhodes, 1994). This issue seems to warrant particular attention for those youth who are already exhibiting some degree of personal vulnerability.

Limitations and Directions for Future Research

Several limitations of the present review also are noteworthy and should be addressed in future research on youth mentoring programs. One significant issue to be kept in mind is that findings do not necessarily reflect causal effects of either mentoring or the different moderator variables examined. Positive effects of programs are evident in studies using the most well-controlled designs (i.e., random assignment) and in those in which mentoring has been provided alone rather than in combination with other types of intervention. Yet, even these types of investigations clearly are not immune to extraneous sources of influence. Consider, for example, the potential that exists for demand characteristics associated with a youth's involvement in a mentoring program to introduce bias into the responses that youth, parents, and other informants (e.g., teachers) provide on outcome measures. To fully address this particular concern, it will be important for future evaluations to more often incorporate "nonreactive" me asures into their assessments of youth outcomes (e.g., archival records of arrests, educational accomplishments, etc.). Given the increasing prevalence of mentoring as an intervention, the possibility that significant numbers of youth within control groups may themselves be involved in a formal mentoring relationship through involvement in other programs or services also merits greater attention than it appears to have thus far received in evaluation studies.

Inferences regarding the influence of different moderator variables are even more tentative because of the inherently correlational nature of any associations that are found between study characteristics and outcomes within the framework of meta-analysis (see Cooper, 1998, for further discussion). Accordingly, priority should be given to more controlled investigation of the factors identified (e.g., at-risk status) within the context of individual studies in future research. There also clearly is a related need for evaluations to more consistently assess characteristics of the relationships that are actually developed between mentors and youth in programs as a source of influence on outcomes. These types of efforts, furthermore, should be complemented by more in-depth consideration of the wide-ranging circumstances within which mentoring may occur in the life of any given youth.

Issues relating to the generalizability of findings also are a significant concern. These include possible limitations in the extent to which results can be extrapolated to the much broader range of mentoring programs not included in the present review. The importance of this consideration is underscored by the lack of complete robustness of findings when conducting analyses under the assumption of a random-

rather than fixed-effect model and by the potential for programs that have not received formal evaluation to differ systematically from those that have been subjected to this type of scrutiny. The ability to make predictions about the efficacy of youth mentoring programs in the future is similarly prone to uncertainty given the still evolving status of approaches to intervention in this area. Subsequent programs, for example, may include significant innovations influencing effectiveness that are not reflected in those programs that have received formal evaluation to date. Estimates of effect size derived along basic dimensions of intervention design and evaluation also vary to some extent and thus serve to illustrate specific areas in which conclusions regarding the effectiveness of mentoring programs for youth may require qualification. These include potential liabilities associated with restricting mentoring activities to the school setting, evidence of a relatively weak impact on emotional/psychological outcomes, and, perhaps most notably, absence of compelling support for inferring benefits to youth that extend substantially beyond the end of program involvement. Cumulatively, the preceding considerations strengthen the rationale for ongoing evaluation of youth mentoring programs, especially with respect to those areas for which effectiveness currently is less well established.

A final recommendation is pragmatic in nature. Because of the diversity of published and unpublished sources in which mentoring program evaluations have appeared, a great deal of time and effort was required to locate and obtain the studies included in the present review. Many of these reports predate earlier reviews, but were not included in them perhaps at least in part because of similar practical considerations. To facilitate a more orderly and efficient compilation of mentoring program evaluation data in the future, it is recommended that a research register be created listing all relevant projects that are either in progress or completed. The availability of a research register has proven helpful in other fields of inquiry (Dickersin, 1994) and in the mentoring literature could serve a complementary function to the national data base of programs already in existence (Save the Children, 1999). Integration of research and practice through such mechanisms offers the best prospect for future development, ev aluation, and dissemination of effective mentoring programs for youth.

Table I.

Stem-and-Leaf Plot of Average Effect Sizes for Evaluations of Youth Mentoring Programs (N = 59 Independent Samples)

Stem	Leaf
+1.0	9
+0.9	4
+0.8	3
+0.7	7
+0.6	3479
+0.5	27
+0.4	134689
+0.3	035679
+0.2	2444
+0.1	255667788889
+0.0	01333445567778
-0.0	8
-0.1	
-0.2	7621
-0.3	8
-0.4	
-0.5	1

Table II.

Moderators of Mentoring Program Evaluation Outcomes

Moderator	k	Fixed effects [Q.sub.b]
Report information		
Year of report (a)	59	3.27 (+)
Prior to 1990	25	
1990 or after	34	
Type of report	59	1.60
Unpublished	39	