American Journal of Community Psychology, Vol. 30, No. 2, April 2002 (© 2002)

Effectiveness of Mentoring Programs for Youth: A Meta-Analytic Review

David L. DuBois,[1] Bruce E. Holloway, Jeffrey C. Valentine, and Harris Cooper

University of Missouri at Columbia

We used meta-analysis to review 55 evaluations of the effects of mentoring programs on youth. Overall, findings provide evidence of only a modest or small benefit of program participation for the average youth. Program effects are enhanced significantly, however, when greater numbers of both theory-based and empirically based "best practices" are utilized and when strong relationships are formed between mentors and youth. Youth from backgrounds of environmental risk and disadvantage appear most likely to benefit from participation in mentoring programs. Outcomes for youth at-risk due to personal vulnerabilities have varied substantially in relation to program characteristics, with a noteworthy potential evident for poorly implemented programs to actually have an adverse effect on such youth. Recommendations include greater adherence to guidelines for the design and implementation of effective mentoring programs as well as more in-depth assessment of relationship and contextual factors in the evaluation of programs.

KEY WORDS: youth mentoring; program evaluation; primary prevention; children; adolescents.

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

[1]To whom correspondence should be addressed at 210 McAlester Hall, Department of Psychology, University of Missouri at Columbia, Columbia, Missouri 65211; e-mail: DuBoisD@missouri.edu.

0091-0562/02/0400-0157/0

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 influence 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 discussion in the literature (Darling et al., 1994; Freedman, 1992; Hamilton & Hamilton, 1992; Tierney et al., 1995), with arguments offered in favor of each type of approach.

Further considerations pertain to the procedures used for recruiting prospective mentors and the levels of training and supervision that are provided to mentors once selected (Rhodes, 1994). Background checks and other screening procedures (e.g., interviews) have been included consistently in recommended guidelines for the selection of mentors in programs (Freedman, 1992; National Mentoring Working Group, 1991; Saito & Blyth, 1992). Some programs also have specifically sought out individuals whose backgrounds (e.g., teacher) may make them especially well-suited to forming effective mentoring relationships with youth. There has been less consensus regarding needs for training and ongoing supervision of mentors (Rhodes, 1994) and accordingly programs have varied considerably in these areas. Nevertheless, there is general agreement that some type of orientation should be provided and that mentors should have ongoing support available to them (Freedman, 1992; Hamilton & Hamilton, 1992; National Mentoring Working Group, 1991). Additional recommendations include matching of youth with mentors on the basis of criteria such as gender, race/ethnicity, or mutual interests; communication of guidelines and expectations regarding frequency of mentor–mentee contact and duration of relationships; monitoring fidelity of implementation through mentor logs and other procedures; incorporation of structured opportunities for mentor–mentee interaction; and provisions for the support and involvement of parents (Freedman, 1992; Hamilton & Hamilton, 1992; National Mentoring Working Group, 1991; Saito & Blyth, 1992).

Characteristics of Youth

The significance attached to mentoring relationships as a protective influence suggests that programs may provide greater benefits to youth who can be considered "at-risk" by virtue of individual and/or environmental circumstances (Rhodes, 1994). Accordingly, these youth have been the focus of a large proportion of mentoring programs (Freedman, 1992) and currently constitute the majority of all those receiving mentoring (McLearn, Colasanto, & Schoen, 1998). Other specific subgroups that have been targeted by programs include youth from single-parent homes (e.g., BB/BSA) and those belonging to racial or ethnic minority groups (e.g., Royse, 1998). Programs also have been directed toward youth of varying ages and developmental levels. Possible sources of influence on outcomes in this regard include the optimal timing of mentoring as a preventive intervention (Institute of Medicine, 1994) as well as practical issues pertaining to implementation (e.g., receptivity of youth to mentoring at differing stages of development).

Mentor–Mentee Relationships

In order to yield desired outcomes, it may be necessary for programs to establish mentoring relationships between youth and adults that involve patterns of regular contact over a significant period of time (DuBois & Neville, 1997; Freedman, 1992; Slicker & Palmer, 1993). Realization of this aim can be limited, however, in actual practice by difficulties encountered in the recruitment of needed mentors, inadequate levels of mentor–mentee involvement, and premature termination of relationships prior to fulfillment of program expectations (Freedman, 1992; Hamilton & Hamilton, 1992). The extent to which mentoring relationships with consistent and sustained patterns of interaction are actually formed in programs therefore represents a potentially important source of variation in outcomes. A related, methodological consideration is whether youth with relationships that fail to meet criteria for minimum levels of contact or longevity are excluded from analyses of program effectiveness. When this is done the result may be an unduly positive assessment of the benefits that can be realistically expected for all youth referred to a given mentoring program (Grossman & Tierney, 1998).

Assessment of Outcomes

Mentoring programs have been conceptualized as potentially affecting youth in a wide variety of areas, including emotional and behavioral

functioning, academic achievement, and employment or career development. One important concern therefore is whether benefits of mentoring are evident across this diverse range of proposed outcomes. Further considerations include the type of data source or informant utilized as well as the timing of outcomes assessment relative to the active period of program operation. To the extent that effects on youth are evident both across multiple sources of data and at follow-up assessments, this would provide particularly strong support for the effectiveness of mentoring programs.

This Study

The specific aims of this study are two-fold: (a) to objectively assess the overall effects of mentoring programs on youth and (b) to investigate possible variation in program impact in association with factors relating to each of the aforementioned areas (i.e., program design and implementation, youth characteristics, mentor–mentee relationships, and assessment of outcomes). The primary goal of the latter analyses is to help identify promising directions for enhancing program effectiveness. Both theory-based and empirically based indices of best practices for mentoring interventions are developed for use in this portion of the research. These indices are utilized in an effort to identify specific constellations of program characteristics associated with enhanced effectiveness. Because of the importance attributed to relationship factors as moderators of program outcomes, supplementary analyses also are conducted of comparisons that have been made in several studies within the intervention group on the basis of relevant features or characteristics of the relationships formed between mentors and youth.

METHOD

Literature Search Procedures

Three primary methods were used to locate evaluations of youth mentoring programs. First, computer searches of PsychINFO, ERIC, Medline, and Dissertation Abstracts reference databases were run using both subject terms (e.g., mentor) and textwords (e.g., Big Brother) to identify relevant articles. The time frame for each search was from 1970, when research on the outcomes of mentoring programs began to appear, through 1998. Second, in an effort to further identify possible unpublished evaluation studies, a search of the Internet was conducted using several search engines (e.g., Yahoo!).

Finally, the reference sections of reports that met criteria for inclusion in the meta-analysis were examined to determine if reference was made to other potentially relevant reports.

Criteria for Including Studies

To be included in the present review, studies needed to meet several criteria. First, the program evaluated needed to involve mentoring as the practice has been defined commonly in the literature (Nettles, 1991; Rhodes, 1994). To maintain consistency with the prevailing view of mentoring as entailing a one-on-one relationship, programs in which mentoring appeared to have occurred primarily on a small group basis were not included. Similarly, because mentoring generally has been regarded as referring to a relationship between an older, more experienced mentor and a younger protegee (Rhodes, 1994), peer tutoring or mentoring programs were excluded from the present review, although those in which older youth (e.g., teenagers) served in a mentor capacity for younger children were eligible for inclusion. Also excluded were those programs in which the adults involved in forming relationships with youth were mental health professionals (e.g., social workers).[2] Second, the study had to examine empirically the effects of participation in a mentoring program, either by preprogram versus postprogram comparison on the same group of youth or a comparison between one group of youth receiving mentoring and another group not receiving mentoring drawn from the same population. The decision to include evaluations reporting either of the two types of comparisons was based on a desire to increase the number of studies available for review and hence enhance power in tests of both overall effects of mentoring programs and possible moderators of their effectiveness. Finally, the sample used in the evaluation of the program needed to include youth with a mean age of less than 19. A decision also was made to exclude from the review evaluations of two well-established prevention programs that have included mentoring-related components in their design: the Adolescent Diversion Project (Davidson & Redner, 1988) and the Primary Mental Health Project (Cowen et al., 1996). Neither of these programs focuses specifically on mentoring and both already have received extensive consideration in prior literature reviews.

[2]The criterion relating to use of mental health professionals as mentors resulted in the well-known Cambridge–Somerville study being excluded from consideration because social workers served as the companions to youth in this intervention (see McCord, 1992). The findings of this intervention frequently are cited as an example of the potential for preventive interventions to have unanticipated negative effects on participants; as will become apparent, however, a similar lesson can be drawn from the results of those evaluation studies of mentoring programs that did meet criteria for inclusion in the present review.

Search Outcome

Search procedures identified 59 separate research reports that met criteria for inclusion and for which information was available to allow for the computation of one or more effect sizes. One report (Tierney et al., 1995) was simply an earlier, unpublished version of a large scale multisite evaluation of BB/BSA programs that was subsequently published (Grossman & Tierney, 1998). Two articles presented immediate posttest and then 2-year follow-up findings for the same sample of youth participating in the Buddy System mentoring program (Fo & O'Donnell, 1975; O'Donnell, Lydgate, & Fo, 1979) and thus were considered for the purposes of the present review to constitute only one independent study. Finally, two additional articles (Fo & O'Donnell, 1974; Rhodes, Haight, & Briggs, 1999) presented findings on subsamples of youth from the preceding, larger evaluations of BB/BSA and the Buddy System program. These reports were excluded from the present review to avoid overlap in samples across reports. On the basis of the foregoing decisions, a total of 55 independent studies or reports were retained for further analysis.

Effect Size Calculations

Effect sizes were computed as *d*-indexes, or standardized mean differences (Cohen, 1988). The *d*-index expresses the difference between two group means in terms of their common standard deviation. In the present context, *d*-indexes were calculated both for comparisons of preprogram versus postprogram means for a given group of youth participating in a mentoring program as well as for comparisons between one group of youth receiving mentoring and another group not receiving mentoring. If information relevant to both types of comparisons was available, separate *d*-indexes were computed for each form of comparison. Whenever possible, *d*-indexes were calculated from means and standard deviations provided by the report writers. When means and standard deviations were not provided but the values of corresponding statistical tests of mean differences were given, formulas provided by Rosenthal (1994) were used to estimate *d*-indexes. When neither type of information was reported, efforts were made to obtain relevant data from the first author of the report. For a few reports ($n = 7$), the only information available to compute *d*-indexes was test statistics associated with analyses (e.g., analysis of covariance) that controlled statistically for other variables, such as pretest scores on the outcome measure. These reports included several evaluations that were based on relatively large samples of youth and that involved random assignment to intervention and control

conditions. Because of these methodological strengths, a decision was made to include them in the review by estimating *d*-indexes on the basis of the test statistics available. Overall, across the 55 independent studies that met criteria for review a total of 575 separate estimates of effect size (i.e., *d*-indexes) were calculated.

Each effect size was weighted by the inverse of its variance to provide more efficient estimation of true population effects (Hedges & Olkin, 1985). This procedure gives greater weight to samples based on larger samples and is the generally preferred alternative (Cooper, 1998). Variance estimates for one-group preprogram versus postprogram effect sizes were based on the formula $v = [1 + (d^2/2)]/n$ (Cooper, Charlton, Valentine, & Muhlenbruck, 2000). Variance estimates for two-group comparisons were calculated using formulas given in Cooper (1998). For purposes of comparison, findings for both unweighted and weighted effect sizes are reported in analyses of overall program effects. Only weighted *d*-indexes are analyzed and reported in analyses of moderator variables. All effect sizes were coded so that positive values indicated differences in directions consistent with a favorable effect of the mentoring program.

Coding of Studies

Each report was coded on multiple characteristics. The characteristics could be divided into six major categories: (a) report information (year of report, published/unpublished); (b) evaluation methodology (type of research design, use of statistical control variables, internal vs. external evaluation, sample size, exclusion of nonactive relationships from analyses); (c) program features (mentoring alone vs. mentoring as part of multicomponent intervention, BB/BSA vs. non-BB/BSA, program goal, geographic location, setting in which mentoring activities occurred, compensation of mentors, monitoring of implementation, characteristics of mentors recruited, procedures for screening prospective mentors, criteria for matching mentors and youth, mentor training, supervision, and support, expectations for frequency of contact and length of mentoring relationships, structured activities for mentors and youth, inclusion of parental support or involvement component); (d) characteristics of participating youth (gender, race/ethnicity, developmental level, single-parent household, socioeconomic background, at-risk status); (e) mentor–mentee relationships (actual frequency of contact, average length); and (f) assessment of outcomes (type of outcome, data source, timing of assessment).[3]

[3] A copy of the coding sheet is available from the first author.

The theory-based index of best practices referred to previously was derived on the basis of the presence of the following 11 program features: monitoring of program implementation, screening of prospective mentors, matching of mentors and youth on the basis of one or more relevant criteria, both prematch and ongoing training, supervision, support group for mentors, structured activities for mentors and youth, parent support or involvement component, and expectations for both frequency of contact and length of relationships. Each of these program features has been included in previous recommendations for establishing effective mentoring programs (Freedman, 1992; Hamilton & Hamilton, 1992; National Mentoring Working Group, 1991; Saito & Blyth, 1992). The number of practices reported for any given program served as its score on this index. The empirically based index of best practices was derived in a similar manner, but was based on those program features that reached or approached ($p < .10$) statistical significance as moderators of effect size in the present investigation. The program features eligible to contribute to this index included those comprising the theory-based index as well as all other aspects of program design and implementation that were examined as potential moderators (e.g., compensation of mentors). It should be noted that the focus on significant individual moderators of effect size in constructing the empirically based index did not necessitate that a trend toward enhanced outcomes would be evident in association with a greater overall number of the features involved being characteristic of particular programs. For such a relationship to be found, the different program features would need to exhibit, to a substantive extent, a nonoverlapping and hence incremental pattern of association with estimates of effect size.

Unit of Analysis

For the present investigation, the independent sample was the primary unit of analysis. Because effect size information was reported for the overall sample in most reports, each report or study generally contributed one independent sample to the analysis. If a study only reported findings separately for different, nonoverlapping subgroups, however, such as boys and girls, it contributed more than one sample to the analysis. Overall, the 55 reports yielded a total of 60 independent samples for analysis.

Within this general framework, a shifting unit of analysis approach was used for determining what constituted an independent estimate of effect (Cooper, 1998). In this procedure, each effect size is first coded as if it were an independent estimate of the intervention's impact. For example, if data for a single report permitted comparison of preprogram and postprogram

scores for both problem behavior and school performance, two separate *d*-indexes would be calculated. When estimating the overall effect of mentoring, these two *d*-indexes would be averaged prior to entry into the analysis, so that the sample contributed only one effect size. However, in an analysis that examined the effect of mentoring on problem behavior and school performance separately, the sample would contribute one effect size to estimates of mean effect size for each of the two relevant categories of outcome measure.

Fixed and Random Effects

A final consideration prior to conducting analyses was the need to decide whether to conceptualize the effect of youth mentoring programs as fixed or random (Hedges & Vevea, 1998). In a fixed-effect analysis, each effect size's variance is assumed to reflect only sampling error (i.e., error solely due to participant differences). This source of error can be taken into account through procedures described previously for weighting effect sizes by sample size. When a random-effect analysis is carried out, a study-level variance component is assumed to be present as an additional source of random influence. Hedges and Vevea (1998) state that fixed-effect models of error are most appropriate when the goal of the research is "to make inferences only about the effect size parameters in the set of studies that are observed (or a set of studies identical to the observed studies except for uncertainty associated with the sampling of subjects)" (p. 3). In general, a random-effect analysis is more conservative because of the consideration of study-level variance as an additional component of error (Wang & Bushman, 1999).

The appropriateness of a random effects model in the present context is suggested both by (a) the large variation that is evident in the implementation and design characteristics of youth mentoring programs (and hence the potential for these factors to constitute significant sources of random error even after taking into account variance associated with specified moderating variables) and (b) interest in drawing inferences about all youth mentoring programs, not just those included in the present review. Alternatively, a fixed effects could be argued to be appropriate to the extent that the effectiveness of those programs that have been subjected to evaluation is of particular interest. Relevant considerations in this regard include the relatively widespread dissemination that some of the most frequently evaluated programs have received (e.g., Big Brothers/Big Sisters) as well as the possibility that programs undergoing formal evaluation may tend to be most representative

of "best practices" in youth mentoring (e.g., more innovative designs, greater monitoring and fidelity in program implementation, etc.). A further statistical consideration is that in the search for moderators, fixed-effect models may seriously underestimate and random-effects models seriously overestimate error variance when their assumptions are violated (Overton, 1998). In view of the foregoing competing sets of concerns, a decision was made to apply both models in all primary study analyses (Cooper et al., 2000). Specifically, all analyses were conducted twice, once employing fixed-effect assumptions and once using random-effect assumptions. This allowed similarities and differences in results across the two types of analyses then to be incorporated into the interpretation and discussion of findings. Both fixed-effect and random-effect analyses were carried out using SAS Software in accordance with procedures described by Wang and Bushman (1999).

RESULTS

Preliminary Analyses

Inspection of the distribution of the 575 unweighted *d*-indexes using a stem-and-leaf plot revealed 11 positive *d*-index values that were more than three interquartile ranges beyond the 75th percentile and thus qualified as statistical outliers according to Tukey's definition (Tukey, 1977). In addition, two effect sizes met a corresponding criterion as a negative outlier. Further investigation revealed that all but three of the positive outlier effect sizes were derived from reports with unusually small samples (i.e., 15 or fewer youth). Of the remaining three effect sizes, one came from a study that examined whether providing a mentor to youth who had attended a summer leadership institute facilitated their completion of a community service project during the following school year (Mertens, 1988). The outcome measured in this study, completion of the community service project, was thus an immediate objective of the program itself. The other two positive effect sizes appeared as relatively isolated findings within the reports involved. In summary, although studies that report unusually large effect sizes merit careful consideration as possible examples of exemplary or "best" practice (Cooper et al., 2000), there was little evidence to support this interpretation in the present review. As a safeguard against these extreme values having undue influence on the findings of subsequent analyses, the effect sizes involved (and two other positive *d*-indexes that approached criterion as outliers) were Winsorized by setting their values to 1.25 or −1.25 in the case

of the negative *d*-index outliers. Because the outlier *d*-index derived from Mertens (1988) was the only effect size available from this report, the study continued to represent an extreme observation at the level of independent sample analysis even after the *d*-index involved had been Winsorized. On the basis of this result and the idiosyncratic features of the outcome assessment involved in this report it was omitted from all subsequent analyses.

The distribution of sample sizes also was examined for extreme values. Nine independent samples, each appearing in separate reports, met or approached criteria as statistical outliers. These samples ranged in size from 373 to 47,775 and included those associated with several, recent large-scale evaluations of mentoring programs such as the national, multisite evaluation of the effectiveness of BB/BSA (Grossman & Tierney, 1998; Tierney et al., 1995) in which 959 youth participated. The largest sample size was accounted for by an evaluation of the Cincinnati Youth Collaborative Mentoring Program (Bruce & Mueller, 1994) in which all youth without mentors in the participating schools served as a comparison group ($n = 46{,}732$) for youth who did receive mentoring ($n = 1{,}043$). Because the procedure for weighting effect sizes was based on sample size, the potential existed for these unusually large samples to have an overwhelming influence on findings. For this reason, all nine sample sizes identified as potential outliers were Winsorized by setting their values to 300.

Overall Effect of Youth Mentoring Programs

After Winsorizing the effect sizes, the average unweighted *d*-index for the 574 effect size estimates included for subsequent analysis was $d = .18$. Using the 59 independent samples involved as the unit of analysis, the average unweighted *d*-index was $d = .23$. The median effect size was $d = .18$. Effect sizes then were evaluated after weighting them by the inverse of their variance, a procedure that involves differing estimation procedures depending on whether a fixed-effects or random-effects model is assumed. Under the fixed-effects assumption, the average effect size for the 59 independent samples was $d = .14$. Thus, making no distinctions among effects based on methodology or program, youth, relationship, or measurement characteristics, the average youth participating in one of the mentoring programs included in the present review scored approximately one eighth of a standard deviation higher in a favorable direction on outcome measures than did the average youth before or without participation in one of these programs. The 95% confidence interval for the weighted *d*-index under the assumption of fixed effects encompassed a lower value of $d = .10$ and an upper value

of $d = .18$. The practical significance of an effect size also can be expressed by describing how outcomes for intervention and control groups overlap (Cooper, 1998, Table 5.3). Using this approach, the average weighted effect size of $d = .14$ under the assumption of fixed effects indicates that across different types of programs, the outcome for the average participant in a youth mentoring program surpassed that of approximately 55% of those in the control group (i.e., the average youth before or without program participation). Under the assumption of random effects, the average weighted effect size estimate increased to $d = .18$, but encompassed a larger confidence interval ranging from .11 to .25.

As a check on the robustness of the preceding findings to the "file-drawer" problem (i.e., lack of publication of studies finding null results), Fail-Safe-N (FSN) calculations were made (Cooper, 1998). The FSN corresponds to the number of null effects that would have to exist in studies not included in the meta-analysis to overturn the conclusion that a significant effect is present. Rosenthal (1979) suggested that FSN be equal to or larger than five times the number of retrieved studies (or, in the present context, independent samples) plus 10. Using $\alpha = .05$ (two-tailed), the FSN for the present study is 513, a value that substantially exceeds the recommended resistance number of 305 ($59 \times 5 + 10$). This result, in combination with the effort that was made to retrieve and include as many unpublished reports as possible in the review, provides confidence that the overall findings consistent with a positive effect of mentoring programs would not be invalidated even in the presence of a significant publication bias in the literature.

A stem-and-leaf display of average *d*-indexes for the 59 independent samples after individual effect sizes had been Winsorized revealed that 51 of the 59 *d*-indexes were in the direction of positive effects for youth mentoring programs (see Table I). Of the remaining eight *d*-indexes, seven were in a negative direction and one corresponded exactly to 0. In each of the former seven cases the negative findings represented half or more of the findings for independent samples within the evaluation. Illustratively, an evaluation conducted by the New York City Board of Education (1986) found declines on measures of school attendance (*d*s of $-.07$ and $-.25$), number of courses passed ($d = -.93$), and grade point average ($d = -.78$) for 79 high school students who received the mentoring component of a multicomponent dropout prevention program. Similarly, youth without a prior major arrest who participated in the Buddy System program referred to previously were found to have higher arrest rates in comparison to youngsters in a randomly assigned control condition (*d*s $= -.27$ and $-.15$ at posttest and 2-year follow-up, respectively), although a trend in the opposite direction favoring program youth was found among youth who had been arrested prior

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

to program involvement (ds = .55 and .43, respectively; Fo & O'Donnell, 1975; O'Donnell et al., 1979).

Moderator Analyses of Mentoring Program Effects

Possible moderators of mentoring program effects were investigated using homogeneity analyses (Cooper & Hedges, 1994; Hedges & Olkin, 1985). This procedure compares the amount of variance in an observed set of effect sizes with the amount of variance that would be expected by sampling error alone (as well as other sources of random influence when assuming a random effects model); the homogeneity statistic for this type of comparison is referred to commonly as Q_b and it follows a chi-square distribution. In reporting these statistics in this study, both degrees of freedom and the number of samples involved, k, will be noted. Whenever feasible, the significance of a potential moderator was tested with the moderator treated as a continuous variable in the homogeneity analysis. This approach was designed to maximize sensitivity in the detection of relevant effects. In several instances, however, it was necessary to treat moderators as categorical variables in analyses because of their inherently categorical nature (e.g., type of outcome measure) or because the degree of variation that was observed across potential values of the variable was not sufficient to justify treatment

as a continuous variable. To facilitate interpretation of results (Cooper et al., 2000), in instances in which moderators were tested as continuous variables average effect sizes are reported for two or more discrete ranges of values of the variable involved.

An examination of frequency distributions for the coded variables revealed some factors for which there was little or no variation discernible across studies or samples. Illustratively, with regard to research design, it was found that in reports comparing two groups of youth (i.e., those receiving mentoring and those not receiving mentoring) that the youth in the control or comparison group typically received no intervention ($n = 37$) as opposed to some type of intervention other than mentoring ($n = 4$). Similarly, the age level of mentors when able to be coded was predominantly in the early adulthood range (19–29 years old; $n = 12$), with a small minority ($n = 6$) in middle adulthood (30–54) and one in late adulthood (55 or older). Lack of variation to this extent effectively prohibited reliable analysis of the factors involved as possible moderators of mentor program effects. There also were several variables for which certain categories had to be combined in order to obtain distributions and numbers of categories that would be suitable for the purpose of moderator analyses. For example, although 44 different types of outcomes had been distinguished in the original coding, these were collapsed into five more general categories of measures for moderator analyses (i.e., emotional/psychological well-being, problem or high-risk behavior, social competence, academic/educational, and career/employment). Finally, in some instances the information required to code the variable was reported for only a minority of samples. Of particular note, information regarding the frequency with which mentors actually had contact with youth in programs and the amount of time that relationships lasted each were reported for only 12 of the 59 independent samples. Because of the theoretical importance of these and other selected moderators, they were nonetheless retained for analysis; clearly, however, the results obtained must be regarded as highly tentative and exploratory in nature.

Before investigating individual moderators of effect sizes, it is important to conduct a homogeneity analysis to test whether there is variability in effect sizes greater than that which would be expected by sampling error around a single population value (Cooper, 1998). The results of this analysis suggested that the *d*-indexes were not all estimating the same underlying population value, $Q(58, k = 59) = 227.70$, $p < .001$, and thus that it was appropriate to look for characteristics potentially involved in moderating effect size. In view of the relatively small number of independent samples available for analysis and limited variability across levels of several potential moderator variables, those findings that approach but do not reach a conventional level of statistical significance (i.e., ps $< .10$) are reported in all tests of individual

moderators. Caution in the interpretation of findings pertaining to the significance of individual moderators is nonetheless certainly warranted given the large number of tests involved and the associated potential for Type I error.

Report Information

As shown in Table II, neither year of report nor whether the report was published (i.e., in a journal article or book) or unpublished (e.g., dissertation) were significant moderators of effect size. Under the assumption of fixed effects, however, year of report did approach significance as a moderator, $Q(1, k = 59) = 3.27$, $p < .10$, with a trend evident toward larger effect sizes for more recent studies (see Table II).

Evaluation Methodology

Neither type of control (pretest–posttest vs. posttest–posttest comparison) nor type of two-group design (random assignment vs. nonequivalent group) were significant moderators of effect size. However, among studies employing nonequivalent two-group designs, there was a trend under the assumption of fixed effects for those that made some attempt to match youth to report larger effects ($d = .20$) than those that did not ($d = .02$), $Q(1, k = 26) = 3.41$, $p < .10$. Among the remaining methodological characteristics of studies that were examined, only sample size was a significant moderator of effect size. This relation was evident for both fixed effects, $Q(1, k = 59) = 8.90$, $p < .01$, and random effects, $Q(1, k = 59) = 8.07$, $p < .01$. As shown in Table II, this finding reflected a tendency for studies with smaller sample sizes to report larger program effects.

When investigating possible substantive moderators of effect size in a meta-analysis, it is recommended that the influence of relevant methodological factors be controlled for statistically (Durlak & Lipsey, 1991). In the present context, these factors included sample size and whether matching was used for nonequivalent groups, given that each exhibited a significant or nearly significant association with effect size. As noted, type of control used as the basis for deriving effect size estimates (i.e., single group pretest–posttest comparison or two group posttest–posttest comparison) was not found to be a significant moderator of effect size. Nevertheless, because of the statistical differences involved in derivation and weighting of each type of effect size estimate, it was desirable to control for any residual variation associated with this methodological factor as a possible source of influence on findings. To implement statistical control for the methodological

Table II. Moderators of Mentoring Program Evaluation Outcomes

		Fixed effects			Random effects		
Moderator	*k*	Q_b	*d*	±95% CI	Q_b	*d*	±95% CI
Report information							
Year of report[a]	59	3.27†			1.07		
Prior to 1990	25		.10	.08		.16	.11
1990 or after	34		.16	.05		.19	.05
Type of report	59	1.60			0.19		
Unpublished	39		.16	.06		.20	.09
Published	20		.10	.07		.16	.12
Evaluation methodology							
Type of control[b]	74	0.19			0.00		
Pretest–posttest	33		.13	.05		.18	.09
Posttest–posttest	41		.14	.06		.18	.09
Type of two-group design	41	0.52			0.64		
Random assignment	15		.12	.10		.12	.11
Nonequivalent group	26		.16	.08		.18	.09
Was nonequivalent group matched?	26	3.41†			0.97		
No	6		.02	.17		.10	.22
Yes	20		.20	.09		.22	.12
How was *d* derived?	59	0.07			0.47		
Unadjusted means & *SD*s	52		.14	.05		.19	.08
Adjusted means & *SD*s	7		.13	.10		.13	.18
Who did evaluation?	57	2.65			0.39		
Internal	35		.11	.05		.16	.09
External	22		.19	.08		.21	.12
# of youth in analyses[a]	59	8.90**			8.07**		
<65	30		.25	.09		.26	.12
≥65	29		.11	.05		.14	.08
Nonactive relationships	55	1.26			0.07		
Included	24		.11	.07		.17	.11
Excluded	31		.16	.06		.19	.10
Program features[c]							
Type of program	59	0.27			0.00		
Mentoring alone	38		.16	.06		.16	.09
Multicomponent	21		.14	.07		.17	.10
BB/BSA program?	59	0.05			0.24		
Yes	8		.14	.11		.12	.19
No	51		.15	.05		.17	.07
Program goal	59	4.51			3.89		
Psychosocial	21		.14	.07		.16	.11
Instrumental	28		.21	.07		.22	.10
Both	10		.08	.10		.05	.15
Geographic location	49	0.72			0.06		
Large urban	21		.14	.06		.16	.11
Other	28		.19	.08		.18	.11
Setting for mentoring activities[d]	59	7.19†			2.66		
Community	29		.14	.06		.15	.09
School	16		.07	.11		.11	.14
Workplace	6		.24	.17		.24	.22
Other	8		.28	.13		.27	.18

(*Continued*)

Table II. (*Continued*)

Moderator	*k*	Fixed effects			Random effects		
		Q_b	*d*	±95% CI	Q_b	*d*	±95% CI
Mentor compensation	51	0.09			0.02		
No	41		.15	.06		.17	.08
Yes	10		.17	.08		.18	.14
Monitoring of implementation[d,e]	59	5.59*			2.71†		
No[f]	15		.06	.09		.06	.14
Yes	44		.18	.05		.19	.07
Gender of mentors (% male)[a]	28	1.12			0.13		
Majority female	14		.24	.09		.24	.10
Majority male	5		.13	.10		.12	.15
All male	9		.13	.14		.18	.07
Race/ethnicity of mentors (% White)[a]	16	0.24			0.04		
Majority White	10		.21	.09		.22	.14
Majority non-White	6		.21	.13		.22	.19
Mentor background: Helping role/ profession[d]	50	5.75*			3.58†		
No[f]	38		.09	.05		.10	.07
Yes	12		.26	.12		.25	.14
Screening of prospective mentors[e]	59	2.63			0.21		
No[f]	31		.11	.07		.15	.10
Yes	28		.18	.06		.18	.09
Matching of mentors and youth[e]	59	2.35			0.05		
No[f]	26		.11	.07		.16	.10
Yes	33		.18	.06		.17	.09
Mentor–youth matching: Gender	33	0.04			0.07		
No	12		.19	.11		.16	.14
Yes	21		.18	.06		.18	.09
Mentor–youth matching: Race	33	2.47			1.71		
No	26		.15	.06		.15	.08
Yes	7		.26	.11		.26	.15
Mentor–youth matching: Interests	33	0.97			0.57		
No	19		.15	.08		.15	.11
Yes	14		.21	.08		.20	.10
Mentor prematch training[e]	59	0.22			0.17		
No[f]	15		.13	.12		.13	.16
Yes	44		.16	.05		.17	.07
Supervision of mentors[e]	59	1.99			0.55		
No[f]	29		.11	.07		.14	.10
Yes	30		.18	.06		.19	.09
Ongoing training of mentors[d,e]	59	5.58*			4.44*		
No[f]	42		.11	.06		.11	.08
Yes	17		.22	.07		.26	.11
Support groups for mentors[e]	59	2.26			1.25		
No[f]	48		.13	.05		.14	.08
Yes	11		.21	.09		.24	.14
Structured activities for mentors/ youth[d,e]	59	6.36*			5.54*		
No[f]	35		.11	.06		.10	.08
Yes	24		.22	.07		.25	.10

(*Continued*)

Table II. (*Continued*)

Moderator	*k*	Fixed effects Q_b	Fixed effects *d*	Fixed effects ±95% CI	Random effects Q_b	Random effects *d*	Random effects ±95% CI
Parent support/involvement[d,e]	59	9.18**			3.44†		
No[f]	46		.11	.05		.13	.07
Yes	13		.27	.09		.27	.13
Expectations: Frequency of contact[d,e]	59	3.92*			0.26		
No[f]	18		.08	.08		.14	.12
Yes	41		.18	.05		.18	.08
Expectations: Length of relationship[e]	59	0.02			0.16		
No[f]	8		.14	.15		.13	.20
Yes	51		.15	.05		.17	.07
Frequency of contact expected[a]	41	1.37			0.36		
≤2 hr per week	19		.15	.09		.16	.10
>2 hr per week	22		.20	.07		.19	.09
Length of relationship expected[a]	51	1.66			0.35		
<12 months	34		.16	.06		.18	.10
≥12 months	17		.14	.07		.16	.12
Best practices: Theory-based[a]	59	12.48***			4.54*		
<6	27		.04	.08		.07	.10
≥6	32		.20	.05		.22	.08
Best practices: Empirically based[a]	59	20.51***			13.65***		
<4	36		.08	.06		.09	.09
≥4	23		.24	.07		.25	.09
Characteristics of youth[c,g]							
Gender (% male)[a]	42	0.03			0.02		
Majority female	22		.18	.07		.18	.08
Majority male	20		.14	.07		.14	.08
Race/ethnicity (% White)[a]	41	0.19			0.21		
Majority White	15		.14	.11		.14	.12
Majority non-White	26		.19	.06		.19	.07
Developmental level	41	0.53			0.06		
Late childhood/early adolescence	20		.17	.06		.17	.11
Middle/late adolescence	21		.13	.09		.15	.12
Single-parent family	59	1.61			0.31		
No[f]	47		.13	.05		.14	.07
Yes	12		.20	.10		.18	.14
Low socioeconomic status	59	2.82†			1.07		
No[f]	37		.11	.06		.12	.08
Yes	22		.19	.07		.19	.09
At-risk status	55	16.20**			8.73*		
None	5		.14	.12		.15	.18
Individual	21		.00	.08		.03	.10
Environmental	18		.18	.07		.17	.10
Both	11		.25	.10		.26	.13
Mentor–mentee relationships[c,g]							
Average frequency of contact[a]	12	0.08			0.08		
<3 hr per week	7		.17	.13		.17	.17
≥3 hr per week	5		.20	.09		.20	.16

(*Continued*)

Table II. (*Continued*)

Moderator	*k*	Fixed effects			Random effects		
		Q_b	*d*	±95% CI	Q_b	*d*	±95% CI
Average length of relationship[a]	12	0.05			0.05		
<1 year	6		.23	.11		.23	.11
≥1 year	6		.14	.13		.14	.13
Assessment of outcomes[c,g]							
Type of outcome[b]	99	4.22			3.19		
Emotional/psychological	20		.09	.08		.10	.12
Problem/high-risk behavior	15		.19	.07		.21	.12
Social competence	11		.16	.09		.15	.15
Academic/educational	43		.13	.05		.11	.08
Career/employment	10		.19	.12		.22	.16
Data source[b]	82	6.15			1.21		
Youth	35		.18	.06		.18	.11
Parent	8		.16	.14		.22	.24
Teacher	7		.25	.14		.21	.23
Administrative records	32		.10	.06		.11	.11
Timing of assessment[b]	74	0.66			0.32		
During program	24		.12	.07		.13	.10
Immediate posttest	39		.14	.06		.16	.08
Follow-up	11		.10	.09		.12	.14
Length of follow-up[a]	11	0.14			0.02		
≤1 year	5		.11	.11		.10	.13
>1 year	6		.09	.13		.10	.15

[a]This variable was utilized as a continuous variable in moderator analyses.
[b]Individual samples in some instances contributed effect sizes to more than one category of this potential moderator variable (see discussion of "Unit of Analysis" in text for details); for this reason, the overall value of *k* for the test of this variable as a moderator is greater than the total number of independent samples included in the review (i.e., 59).
[c]Analyses of this category of moderator variables includes statistical control for the following methodological factors: sample size, type of control (i.e. pretest–posttest vs. posttest–posttest), and whether matching was used for nonequivalent groups.
[d]This program feature was included in the empirically based index of best practices.
[e]This program feature was included in the theory-based index of best practices.
[f]Includes samples from reports in which the moderator was not mentioned.
[g]Analyses of this category of moderator variables includes statistical control for theory-based and empirically based indices of best practices.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

factors, all *d*-index estimates were residualized on sample size, whether matching was used for nonequivalent groups, and type of control/effect size estimate (see Cooper et al., 2000, for details). The resulting adjusted *d*-indexes were used in all subsequent moderator analyses.[4]

[4]To investigate the extent to which findings of the remaining moderator analyses were affected by introducing control for relevant methodological features of studies, supplementary analyses were conducted in which this type of control was not included (i.e., all moderator analyses reported in Table II relating to program features, characteristics of youth, mentor–mentee relationships, and assessment of outcomes). Results of these analyses were generally unchanged from those that did include methodological controls with only a limited number

Program Features

As shown in Table II, whether mentoring was provided alone or as part of a multicomponent program was not a significant moderator of effect size. Similarly, neither the comparison of BB/BSA versus non-BB/BSA programs[5] nor the comparison of programs according to psychosocial versus instrumental goals yielded significant differences in effect sizes. Effect size also did not demonstrate a significant relation with geographic program location, the setting in which mentoring activities took place, or compensation of mentors (see Table II). Under the assumption of fixed effects, however, there was a trend indicating setting for mentoring activities as a moderator, $Q(4, k = 59) = 7.19$, $p < .10$, with lower effect sizes for programs that were based in schools ($d = .07$) as opposed to other settings such as the workplace ($d = .24$) or community ($d = .14$). In addition, for the fixed-effect analysis, monitoring of program implementation was a significant moderator of effect size, $Q(1, k = 59) = 5.59$, $p < .05$, with larger effect sizes found for programs that reported use of procedures for monitoring implementation ($d = .18$) in comparison to those that did not ($d = .06$). This moderator also approached significance ($p < .10$) within the random-effect analysis.

With regard to characteristics of mentors, neither gender nor race/ethnicity were significant moderators of effect size. Utilization of mentors with a background in a helping role or profession (e.g., teacher), however,

of exceptions evident in terms of which variables either approached ($p < .10$) or reached ($p < .05$) significance as moderators. Furthermore, these latter variations notwithstanding, the pattern of effect size estimates across relevant categories or levels of each moderator variable without control for methodological factors was found to be substantively similar to that which was obtained in primary analyses when this type of control was included (see Table II).

[5] It will be recalled that the evaluations of BB/BSA programs included a recent large scale, multisite investigation of the effectiveness of BB/BSA (Grossman & Tierney, 1998; Tierney et al., 1995). Using means and standard deviations provided by the study investigators, it was possible for the present review to compute both preprogram versus postprogram effect size estimates (i.e., change for youth participating in BB/BSA) and postprogram effect size estimates (i.e., youth in BB/BSA vs. those in randomized control group at the 18-month follow-up) for all of the 46 outcome measures included in the research. Average pre–post and post–post effect size estimates were .02 and .05, respectively, and thus lower than those evident overall for the eight BB/BSA evaluations included in the present review (*d*s of .14 and .12 under assumptions of fixed and random effects, respectively). This finding is not necessarily consistent with the manner in which results of the large-scale evaluation frequently have been cited by the media (e.g., "Big government," 1995) and others as demonstrating a large impact for mentoring relationships. Several factors may be relevant to consider in this regard, however, including the use of a nonstandard methodology for deriving estimates of the magnitude of program effects in original reports of the research (e.g., Tierney et al., 1995), the equal weight given to all outcome measures in the present analysis as opposed to, for example, only those for which statistically significant effects were found, and, finally, possible enhanced sensitivity to detecting program influences when incorporating statistical control for variations in baseline characteristics of study participants as was done in the primary research (Grossman & Tierney, 1998).

was a significant moderator of effect size under the assumption of fixed-effects, $Q(1, k = 50) = 5.75$, $p < .05$. Evaluations of programs that used these types of mentors reported larger effect sizes ($d = .26$) than those for which utilization of such mentors was not indicated ($d = .09$). This moderator also approached significance ($p < .10$) within the random-effect analysis.

The use of procedures for screening prospective mentors was not related significantly to effect size, nor was matching of mentors and youth on the basis of relevant criteria. Furthermore, among those programs that did utilize matching procedures, different types of criteria for matching (i.e., gender, race/ethnicity, or interests) were not significant moderators of effect size.

Although the provision of initial, prerelationship training to mentors was not related significantly to effect size, a difference was found with regard to the provision of ongoing training during relationships for both fixed effects, $Q(1, k = 59) = 5.58$, $p < .05$, and random effects, $Q(1, k = 59) = 4.44$, $p < .05$. Specifically, those programs in which mentors received ongoing training reported larger effects (*d*s of .22 and .26 for fixed and random effects, respectively) than those in which this type of training was not indicated to have been made available (*d*s of .11). As can be seen in Table II, provision of structured activities for mentors and youth and inclusion of a parent support or involvement component also were significant moderators of effect size under the assumption of fixed effects, $Q(1, k = 59) = 6.36$, $p < .05$, and $Q(1, k = 59) = 9.18$, $p < .01$, respectively; under the assumption of random effects, provision of structured activities remained a significant moderator ($p < .05$) and parent support/involvement approached significance ($p < .10$). As can be seen in Table II, evaluations of programs that included these features reported larger effect sizes than did those of other programs. Supervision and support groups for mentors were not found to be significant moderators of effect size.

The final individual program features examined were expectations for frequency of contact and duration of relationships between mentors and youth. Expectations regarding frequency of contact was a significant moderator of effect size under fixed effects, $Q(1, k = 59) = 3.92$, $p < .05$, with larger effect sizes reported in evaluations of programs that did include this type of expectation ($d = .18$) in comparison to other programs ($d = .08$). This program feature was not a significant moderator, however, when conducting a random-effect analysis. Expectations regarding the duration of relationships was not a significant moderator of effect size for the fixed-effect or random-effect analysis. In addition, among those programs for which either type of expectation was reported, variations in the frequency of contact or length of relationship expected were not found to be related significantly to effect size (see Table II).

The theory-based index of best practices described previously was found to be a significant moderator of effect size both under fixed effects, $Q(1, k = 59) = 12.48$, $p < .001$, and random effects, $Q(1, k = 59) = 4.54$, $p < .05$. As shown in Table II, larger effect sizes were reported in evaluations of programs that engaged in a majority of the 11 proposed best practices (*d*s of .20 and .22 for fixed and random effects, respectively) in comparison to other programs (*d*s of .04 and .07). This indicated cumulative contribution of theory-based "best practice" indicators to the prediction of greater program effect sizes and is consistent with the previously described analyses in which 5 of the 11 indicators involved were found to reach or approach significance as individual moderators of effect size; it is also noteworthy in this regard that although nonsignificant, variation in average effect size estimates for each of the remaining six indicators was in the direction of larger effects in conjunction with the presence of the relevant "best practice" indicator (see Table II). To further ensure that results for the theory-based index of best practices did not reflect the influence of a single isolated program feature, a sensitivity analysis was conducted. Specifically, the index was reexamined as a moderator in a series of analyses that sequentially excluded each of the 11 theory-based indicators of best practices from the index. Under the assumption of fixed effects, the index remained a significant moderator of effect size regardless of which indicator was excluded from consideration (*p*s $< .01$). Similarly, it also generally remained a significant moderator under the assumption of random effects (*p*s $< .05$). The exceptions in the latter instance were that the finding approached significance ($p < .10$) only when removing either ongoing training for mentors or provision of structured activities for mentors/youth from the index.

As described previously, the empirically based index of best practices was based on the program features that reached or approached significance as individual moderators of effect size in the present investigation. A total of seven program features met this criterion (see Table II).[6] For purposes of incorporating setting for mentoring activities into the empirically based best practices index, all settings other than school were considered a "best practice" (i.e., community, workplace, other) given that consistently larger effect sizes were found to be associated with these programs relative to

[6]Two program features (i.e., setting for mentoring activities and expectations for frequency of contact) were significant moderators of effect size in the fixed-effect analysis, but did not reach or approach significance in the random-effect analysis. Nevertheless, to allow for comparability of findings across fixed- and random-effect analyses, a decision was made to compute a single empirically based best practices index for each study that included consideration of the presence or absence of both of the program features involved. This relatively inclusive strategy is consistent with the inherently more conservative nature of random-effect analyses in general (Wang & Bushman, 1999) as well as their potential to obscure significant moderator effects when underlying assumptions are not met (Overton, 1998).

those for which mentoring activities were indicated to have been primarily school-based. The resulting index was found to be a significant moderator of effect size both under fixed effects, $Q(1, k = 59) = 20.51$, $p < .001$, and random effects, $Q(1, k = 59) = 13.65$, $p < .001$. As shown in Table II, evaluations of programs that engaged in a majority of the relevant practices reported substantially larger effect sizes (*d*s of .24 and .25 for fixed and random effects, respectively) than did other programs (*d*s of .08 and .09). A sensitivity analysis conducted using procedures described previously revealed that this moderating effect was not attributable to any single program feature included in the index (*p*s $< .001$ for both fixed and random effects).

The results of the preceding analyses indicate a trend for greater numbers of theory- and empirically based best practice indicators to be associated with larger effect sizes. They do not address, however, which specific sets of features included in the best practice indices were most responsible for these trends. Moderate, but statistically significant zero-order correlations were evident both among the 11 theory-based best practice indicators (absolute mean $r = .18$; range from $-.15$ [supervision of mentors and provision of structured activities for mentors/youth] to .57 [supervision of mentors and matching of mentors and youth]) and among the 7 empirically based indicators (mean $r = .23$; range from $-.54$ [nonschool setting for mentoring activities and background in helping role/profession for mentors] to .37 [monitoring of implementation and background in helping role/profession for mentors]). Accordingly, a multivariate approach was used to investigate the extent to which specific indicators comprising each index made independent contributions to the prediction of effect size. Specifically, a forward selection stepwise multiple regression procedure was used to construct a single best-fitting equation for each type of index. The criterion used for variable entry was a significant or nearly significant ($p < .10$) unique contribution to the prediction of effect size independent of other variables already included in the model; in addition, variables already included as predictors were eligible for removal at successive steps if their contributions no longer approached significance. Two alternatives to the squared multiple correlation (i.e., R^2) have been proposed for quantifying the degree to which a given set of predictors in a fixed-effects regression model explain or account for variation in effect sizes. Hedges (1994) recommended use of a descriptive statistic called the Birge ratio (R_B); this ratio estimates the ratio of between-studies variation in effects to the variation due to (within-study) sampling error. Larger values indicate greater degrees of unexplained variation relative to a model with exact fit (i.e., Birge ratio of 1); a Birge ratio of 1.5, for example, suggests that there is 50% more between-studies variation than might be expected given the within-study sampling variance. Hedges (1994)

as well as Wang and Bushman (1999) also discussed computing the proportion of "explainable" between-study variation in effect size that is accounted for by a set of predictors (i.e., the squared multiple correlation divided by the maximum squared correlation that is possible, taking into account nonsystematic within-study sources of error that are inherently unable to be accounted for by study-level predictors).

For theory-based best practice indicators, the best-fitting regression under the assumption of fixed effects included parent support/involvement ($b = .13$, $p < .05$) and structured activities for mentors/youth ($b = .08$, $p < .10$) as predictors ($R^2 = .12$; $R_B = 1.66$; 27% of "explainable" between-study variation). Under the assumption of random effects, the best-fitting model included structured activities for mentors/youth ($b = .14$, $p < .05$) and ongoing training for mentors ($b = .12$, $p < .10$) as predictors.

For empirically based best practice indicators, the best-fitting model under the assumption of fixed effects included four predictors ($R^2 = .24$; $R_B = 1.48$; 53% of "explainable" between-study variation): nonschool setting for mentoring activities ($b = .26$, $p < .001$), mentor background in a helping role/profession ($b = .22$, $p < .01$), parent support/involvement ($b = .13$, $p < .05$), and structured activities for mentors/youth ($b = .08$, $p < .10$). Under the assumption of random effects, the best-fitting model included structured activities for mentors/youth and ongoing training for mentors as predictors and thus was identical to the model identified in the random-effect analysis for theory-based best practice indicators.

Because of the associations found between program practices and estimates of effect size, it was possible that differences in implementation of relevant practices across studies could introduce bias into analyses of other types of potential moderator variables. To address this concern, all *d*-indexes were adjusted for their associations with both the theory-based and empirically based indices of best practices, using the same regression procedure that had been employed previously to control for methodological factors as a confounding influence. The resulting *d*-indexes, now adjusted for methodological factors and both indices of best practices, were utilized in the remaining moderator analyses.[7]

[7]Supplementary analyses for the remaining categories of moderator variables (i.e., characteristics of youth, mentor–mentee relationships, and assessment of outcomes) also were conducted without the incorporation of control for the theory- and empirically based best practice indices. With only one exception, results did not differ substantively from those obtained in primary analyses that did include control for the best practice indices (see Table II). The exception involved developmental level approaching significance ($p < .06$) as a moderator of effect size under the assumption of fixed effects when removing control for the best practice indices (*d*s of .21 and .10 for late childhood/early adolescence and middle/late adolescence, respectively), whereas it did not do so in primary analyses. This difference appears attributable to greater numbers of both theory- and empirically based best practice indicators for programs in which

Characteristics of Youth

As shown in Table II, the demographic characteristics of youth and their families that were examined (i.e., gender, race/ethnicity, developmental level, single-parent home, socioeconomic status) were not significant moderators of effect size. Under the assumption of fixed effects, however, socioeconomic status did approach significance as a moderator, $Q(1, k = 59) = 2.82$, $p < .10$. Larger effect sizes were reported for samples of youth from primarily low socioeconomic backgrounds ($d = .19$) than for other samples ($d = .11$). Consistent with this trend, at-risk status was found to be a significant moderator of effect size under the assumptions of both fixed effects, $Q(3, k = 55) = 16.20$, $p < .01$, and random effects, $Q(3, k = 55) = 8.73$, $p < .05$. Effect sizes were largest for samples of youth experiencing both individual and environmental risk factors (*d*s of .25 and .26 for fixed and random effects, respectively) or environmental risk alone (*d*s of .18 and .17). Average effect size estimates were somewhat lower for the relatively small number of samples in which youth were not indicated to be experiencing either type of risk (*d*s of .14 and .15), with the associated confidence intervals not consistently allowing for the inference of an overall positive effect of mentoring. Finally, near-zero average estimates of effect size were evident for those samples of youth indicated to be experiencing individual risk factors alone (*d*s of .00 and .03). To further investigate the latter finding, additional analyses were conducted to determine whether program practices were related to the magnitude of indicated effects of mentoring on youth exhibiting individual level risk factors. These analyses sought to examine whether positive effects of mentoring on these youth might be evident for those programs that employed relatively greater numbers of the previously described theory-based and empirically based "best practices." Accordingly, the estimates of effect size utilized for these supplementary analyses were preadjusted only for the previously noted possible methodological confounds (i.e., control for number of "best practices" was removed). The theory-based "best practices" index was found to be a significant moderator of effect size among the 21 independent samples of youth included within the individual risk status category under the assumptions of both fixed effects, $Q(1, k = 21) = 14.81$, $p < .001$, and random effects, $Q(1, k = 21) = 5.35$, $p < .001$. Positive effects of mentoring for these youth were evident when programs engaged in a majority of the relevant practices (*d*s = .20 and .24 for fixed and random effects, respectively, with 95% confidence intervals of

relatively younger youth received mentoring (i.e., *M*s of 6.95 and 4.20 for late childhood/early adolescent and 4.67 and 2.90 for middle/late adolescent, respectively, *p*s $< .01$), such that larger estimates of effect size at earlier stages of development were evident only when failing to control for associated variation in relevant features of the programs involved.

±.13 and .18); by contrast, when this was not the case, average effect size estimates were in a negative direction (ds = −.13 and −.06, confidence intervals of .11 and .18, respectively). A majority of the 21 independent samples involved ($n = 12$) constituted the latter group for which "best practices" were less evident, thus contributing to the lack of an overall positive effect size for this category of risk status. Similar results were obtained when examining the empirically based best practices index as a moderator, for fixed effects: $Q(1, k = 21) = 24.98$, $p < .001$, ds = −.13 and .21, confidence intervals of .11 and .13, for programs engaging in fewer or more than half of the relevant practices, respectively; for random effects, $Q(1, k = 21) = 11.47$, $p < .001$, ds = −.07 and .27, confidence intervals of .17 and .22, respectively, with most samples of youth in the individual-risk status category ($n = 13$) again participating in programs not indicated to be engaging in a majority of the targeted practices.

Mentor–Youth Relationships

As can be seen in Table II, neither reported average frequency of contact between mentors and youth nor length of relationship was a significant moderator of effect size. It will be recalled, however, that these analyses were limited by the small numbers of studies that reported data on either variable.

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, $\chi^2(3) = 11.71$, $p < .01$, with samples of youth in the individual risk status 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 findings 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 relevant 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 95% 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; Hamilton & 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 formidable 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 possibility 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 yet 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 personal 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 relating 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" measures 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, evaluation, and dissemination of effective mentoring programs for youth.

REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis.

*Abbott, D. A., Meredith, W. H., Self-Kelly, R., & Davis, M. E. (1997). The influence of a Big Brothers program on the adjustment of boys in single-parent families. *Journal of Psychology*, *131,* 143–156.

*Abcug, L. (1991). *Teachers Achieving Success With Kids (TASK): A teacher–student mentorship program for at-risk students* (M. S. Practicum, Nova University). (ERIC Document Reproduction Service No. ED330974)

*Aiello, H. S. (1988). Assessment of a mentor program on self-concept and achievement variables of middle school underachievers (Doctoral dissertation, Virginia Polytechnic Institute and State University, 1988). *Dissertation Abstracts International*, *49*(07), 1699A.

*Baldwyn Separate School District, MS. (1982). *A gifted model designed for gifted students in a small, rural high school.* Washington, DC: Office of Elementary and Secondary Education. (ERIC Document Reproduction Service No. ED233847)

*Banta, T. W., & Lawson, S. S. (1980). *Evaluation of the Lenoir City Schools (Tennessee) Retirement Power in Education Project, 1979–1980.* Knoxville, TN: Bureau of Educational Research and Service. (ERIC Document Reproduction Service No. ED201044)

Big government, big brother. (1995, December 25). *The New Republic*, 7.

*Blakely, C. H., Menon, R., & Jones, D. C. (1995). *Project BELONG: Final report.* College Station, TX: Texas A&M University, Public Policy Research Institute.

*Brooks, L. J. (1995). An evaluation of the VCU Mentoring Program (Doctoral thesis, Virginia Commonwealth University, 1995). *Dissertation Abstracts International*, *57*(04), 1481A.

*Bruce, D., & Mueller, E. J. (1994). *Mentoring: A Cincinnati Youth Collaborative program. Evaluation of outcomes 1993–1994.* Cincinnati, OH: Cincinnati Youth Collaborative.

*Buman, B., & Cain, R. (1991). *The impact of short term, work oriented mentoring on the employability of low-income youth.* (Available from Minneapolis Employment and Training Program, 510 Public Service Center, 250 S. 4th St., Minneapolis, MN 55415)

*Cave, G., & Quint, J. (1990). *Career beginnings impact evaluation: Findings from a program for disadvantaged high school students.* New York: Manpower Demonstration Research Corporation. (ERIC Document Reproduction Service No. ED325598)

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

*Colson, S. (1979). Evaluating career education for gifted and talented students. *Journal of Research and Development in Education*, *12,* 51–62.

Cooper, H. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed.). Thousand Oaks, CA: Sage.

Cooper, H., Charlton, K., Valentine, J. C., & Muhlenbruck, L. (2000). Making the most of summer school: A meta-analytic and narrative review. *Monographs of the Society for Research in Child Development*, *65*(1, Serial No. 260).

Cooper, H., & Hedges, L. V. (Eds.). (1994). *Handbook of research synthesis.* New York: Russell Sage Foundation.

Cowen, E. L. (1985). Person-centered approaches to primary prevention in mental health: Situation-focused and competence-enhancement. *American Journal of Community Psychology*, *13,* 31–48.

Cowen, E. L., Hightower, A. D., Pedro-Carroll, J., Work, W. C., Wyman, P. A., & Haffey, W. C. (1996). *School based prevention for at-risk children: The Primary Mental Health Project.* Washington, DC: American Psychological Association.

Darling, N., Hamilton, S. F., & Niego, S. (1994). Adolescents' relations with adults outside the family. In R. Montemayor & G. R. Adams (Eds.), *Personal relationships during adolescence* (pp. 216–235). Thousand Oaks, CA: Sage.

Davidson, W. S., & Redner, R. (1988). The prevention of juvenile delinquency: Diversion from the juvenile justice system. In R. H. Price, E. L. Cowen, R. P. Lorion, & J. Ramos-McKay (Eds.), *Fourteen ounces of prevention: Theory, research, and prevention* (pp. 123–137). New York: Pergamon.

*Davis, H. (1988). A mentor program to assist in increasing academic achievement and attendance of at-risk ninth grade students (Doctoral thesis, University of Pittsburgh, 1988). *Dissertation Abstracts International*, *50*(03), 0580A.

*Davison, A. R. (1994). A culturally-enriched career and tutorial program and the vocational interests and career beliefs of at-risk middle school students (Doctoral dissertation, Lehigh University, 1994). *Dissertation Abstracts International*, *56*(02), 492A.

*Dicken, C., Bryson, R., & Kass, N. (1977). Companionship therapy: A replication of experimental community psychology. *Journal of Consulting and Clinical Psychology*, *4,* 637–642.

Dickersin, K. (1994). Research registers. In H. Cooper & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 71–83). New York: Russell Sage Foundation.

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., & Lipsey, M. W. (1991). A practitioner's guide to meta-analysis. *American Journal of Community Psychology*, *19,* 291–332.

Durlak, J. A., & Wells, A. M. (1997). Primary prevention mental health programs for children and adolescents. *American Journal of Community Psychology*, *25,* 115–152.

Durlak, J. A., & Wells, A. M. (1998). Evaluation of indicated preventive intervention (secondary prevention) for children and adolescents. *American Journal of Community Psychology*, *26,* 775–802.

*Flaherty, B. P. (1985). An experiment in mentoring high school students assigned to basic courses (Doctoral thesis, Boston University, 1985). *Dissertation Abstracts International*, *46*(02), 0352A.

Flaxman, E., Ascher, C., & Harrington, C. (1988). *Youth mentoring: Programs and practices* (Urban Diversity Series No. 97). New York: Teachers College, Columbia University. (ERIC Document Reproduction No. ED308257)

Fo, W. S., & O'Donnell, C. R. (1974). The Buddy System: Relationship and contingency conditions in a community intervention program with nonprofessionals as behavior change agents. *Journal of Consulting and Clinical Psychology*, *42,* 163–169.

*Fo, W. S., & O'Donnell, C. R. (1975). The Buddy System: Effects of community intervention on delinquent offenses. *Behavior Therapy*, *6,* 522–524.

Freedman, M. (1988). *Partners in growth: Elder mentors and at-risk youth.* Philadelphia: Private/Public Ventures.

Freedman, M. (1992). *The kindness of strangers: Reflections on the mentoring movement.* Philadelphia: Public/Private Ventures.

*Galvin, P. F. (1989). Concept mapping for planning and evaluation of a big brother/big sister program. *Evaluation and Program Planning*, *12,* 53–57.

*George, A. Z. (1986). Big Brothers of Greater Los Angeles: An evaluation from a social learning theory approach (Doctoral thesis, California School of Professional Psychology, Los Angeles). *Dissertation Abstracts International*, *46*(08), 2806B.

*Goodman, S. (1972). *Companionship therapy: Studies in structured intimacy.* San Francisco: Jossey-Bass.

*Graber, M. (1985, March). *The effects of an internship program on the psychosocial development of high school students.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

*Green, B. C. (1979). An evaluation of a Big Brothers' program for father-absent boys: An ecobehavioral analysis (Doctoral thesis, New York University, 1979). *Dissertation Abstracts International*, *41*(02), 0671B.

*Grossman, J. B., & Tierney, J. P. (1998). Does mentoring work? An impact study of the Big Brothers Big Sisters program. *Evaluation Review*, *22,* 403–426.

Hamilton, S. F., & Hamilton, M. A. (1992, March). Mentoring programs: Promise and paradox. *Phi Delta Kappan*, 546–550.

*Harmon, M. A. (1995). Reducing drug use among pregnant and parenting teens: A program evaluation and theoretical examination (Doctoral thesis, University of Maryland College Park, 1995). *Dissertation Abstracts International*, *56*(08), 3319A.

*Hayes, G. L. (1998). An evaluation of a staff mentor program for at-risk students in an Oregon high school: CAKE (Caring About Kids Effectively) (Doctoral thesis, Portland State University, 1998). *Dissertation Abstracts International*, *59*(05), 1517A.

*Hayward, B. J., & Tallmadge, G. K. (1995). *Strategies for keeping kids in school: Evaluation of dropout prevention and reentry projects in vocational education* (Final Report). Washington, DC: American Institutes for Research in the Behavioral Sciences. (ERIC Document Reproduction Service No. ED385767)

Hedges, L. V. (1994). Fixed effects models. In H. Cooper & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 285–299). New York: Russell Sage Foundation.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed and random effects models in meta-analysis. *Psychological Methods*, *3,* 486–504.

*Hill, M. J. (1972). *Partners: Community volunteer and probationer in a one-to-one relationship.* Juneau, AL: Department of Health and Social Services, Division of Corrections.

*Howitt, P. S., Moore, E. A., & Gaulier, B. (1998). Winning the battles and the wars: An evaluation of comprehensive, community-based delinquency prevention programming. *Juvenile and Family Court Journal*, *49,* 39–49.

*Huisman, C. (1992). *Student Mentoring Program 1989–1992: Evaluation report.* Portland, OR: Oregon Community Foundation. (ERIC Document Reproduction Service No. ED356701)

Institute of Medicine. (1994). *Reducing risks for mental disorders.* Washington, DC: National Academy Press.

*Johnson, A. W. (1997). Mentoring at-risk youth: A research review and evaluation of the impacts of the Sponsor–A-Scholar Program on student performance (Doctoral dissertation, University of Pennsylvania, 1997). *Dissertation Abstracts International*, *58*(03), 813A.

Johnson, A. W., & Sullivan, J. W. (1995). Mentoring program practices and effectiveness. In M. Galbraith & N. Cohen (Eds.), *Mentoring: New strategies and challenges* (pp. 43–56). San Francisco: Jossey-Bass.

*Joseph, J. A. (1992). *Improving self-esteem of at-risk students.* (Educational Specialist Practicum, Nova University). (ERIC Document Reproduction Service No. ED343106)

*Keenan, J. T. (1992). Evaluation of a Big Brothers/Big Sisters program for children of alcoholics (Doctoral thesis, University of Pennsylvania, 1992). *Dissertation Abstracts International*, *53*(07), 2249A.

*Lakes, K. D. (1997). *The effects of a high school mentoring program on selected academic variables.* Unpublished master's thesis, Wright State University, Dayton, OH.

*Laughrey, M. C. (1990). *The design and implementation of a mentor program to improve the academic achievement of Black male high school students* (Educational Specialist Practicum Report, Nova University). (ERIC Document Reproduction Service No. ED328647)

*Lee, S., Plionis, E., & Luppino, J. (1989). *Keep Youth in School: A community based practice model to keep at risk youth in school: Final report.* Washington, DC: Catholic University of America. (ERIC Document Reproduction Service No. ED314676)

Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research.* Newbury Park, CA: Sage.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, *48,* 1181–1209.

*LoSciuto, L., Rajala, A. K., Townsend, T. N., & Taylor, A. S. (1996). Outcome evaluation of Across Ages: An intergenerational mentoring approach to drug prevention. *Journal of Adolescent Research*, *11,* 116–129.

McCord, J. (1992). The Cambridge–Somerville study: A pioneering longitudinal experimental study of delinquency prevention. In J. McCord & R. J. Tremblay, (Eds.), *Preventing antisocial behavior: Interventions from birth through adolescence* (pp. 196–206). New York: Guilford.

McLearn, K. T., Colasanto, D., & Schoen, C. (1998). *Mentoring makes a difference: Findings from The Commonwealth Fund 1998 Survey of Adults Mentoring Young People.* New York: The Commonwealth Fund.

*McPartland, J. M., & Nettles, S. (1991). Using community adults as advocates or mentors for at-risk middle school students. *American Journal of Education*, *99,* 637–642.

*Mertens, B. L. (1988). An evaluation of the Governors' School Community Leadership Projects and Civic Mentor Network. (Doctoral Dissertation, Seattle University, 1988). *Dissertation Abstracts International*, *49*(08), 2169A.

*Mitchell, H., & Casto, G. (1988). *Team education for adolescent mothers.* Logan, UT: Early Intervention Research Institute. (ERIC Document Reproduction Service No. ED299744)

National Mentoring Working Group. (1991). *Mentoring: Elements of effective practice.* Washington, DC: National Mentoring Partnership.

*Nelson, C., & Valliant, P. M. (1993). Personality dynamics of adolescent boys where the father was absent. *Perceptual and Motor Skills*, *76,* 435–443.

Nettles, S. M. (1991). Community contributions to school outcomes of African-American students. *Education and Urban Society*, *24,* 41–52.

*New York City Board of Education. (1986). *High School Attendance Improvement/ Dropout Prevention Program 1984–1985: Final report* (OEA Evaluation Report). Brooklyn: Office of Educational Assessment. (ERIC Document Service Reproduction No. ED271529)

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

Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*, *3,* 354–379.

Parra, G. R., DuBois, D. L., Neville, H. A., Nanda, S., Mosakowski, K., Sainz, E., et al. (1998, August). *Mentoring relationships for youth: Predictors of perceived benefits and longevity.* Paper presented at the 106th annual convention of the American Psychological Association, San Francisco.

*Powers, L. E., Sowers, J., & Stevens, T. (1995). An exploratory randomized study of the impact of mentoring on the self-efficacy and community-based knowledge of adolescents with severe physical challenges. *Journal of Rehabilitation*, *61*(1), 33–41.

*Powers, S., & McConner, S. (1997). *Project SOAR 1996–1997: Evaluation report.* Tucson, AZ: Creative Research Associates. (ERIC Document Reproduction Service No. ED412269)

*Quint, J. (1991). Project redirection: Making and measuring a difference. *Evaluation and Program Planning*, *14,* 75–86.

*Reller, D. J. (1987). *A longitudinal study of the graduates of the Peninsula Academies: Final report.* Palo Alto, CA: American Institutes for Research in the Behavioral Sciences. (ERIC Document Reproduction Service No. ED327601)

Rhodes, J. E. (1994). Older and wiser: Mentoring relationships in childhood and adolescence. *Journal of Primary Prevention*, *14,* 187–196.

Rhodes, J. E., Haight, W. L., & Briggs, E. C. (1999). The influence of mentoring on the peer relationships of foster youth in relative and non-relative care. *Journal of Research on Adolescence*, *9,* 185–201.

*Rippner, M. (1992). *Using businesses as on-site-schools to increase academic achievement and develop employability skills of at-risk students* (M.S. Assignment, Nova University). (ERIC Document Reproduction Service No. ED352459)

*Roberts, A., & Cotton, L. (1994). Note on assessing a mentor program. *Psychological Reports*, *75,* 1369–1370.

Rosenthal, R. (1979). Replications and their relative utility. *Replications in Social Psychology*, *1,* 15–23.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.

*Rowland, R. G. (1992). An evaluation of the effects of a mentoring program on at-risk students in selected elementary schools in the North East Independent School District (Doctoral dissertation, Texas A&M University, 1991). *Dissertation Abstracts International*, *53*(1), 0039A.

*Royse, D. (1998). Mentoring high-risk minority youth: Evaluation of the Brothers Project. *Adolescence*, *33,* 145–158.

Saito, R. N., & Blyth, D. A. (1992). *Understanding mentoring relationships.* Minneapolis, MN: Search Institute.

Save the Children. (1999). *Overview of Save the Children's National Mentoring Database* [On-line]. Available: www.savethechildren.org/mentors2d.html

*Seidle, F. W. (1982). Big Sisters: An experimental evaluation. *Adolescence*, *17,* 117–128.

*Slicker, E. K., & Palmer, D. J. (1993). Mentoring at-risk high school students: Evaluation of a school-based program. *The School Counselor*, *40,* 327–333.

*Smith, M. A. B. (1990). The teen incentive program: A research and evaluation model for adolescent pregnancy prevention (Doctoral dissertation, Columbia University, 1990). *Dissertation Abstracts International*, *51*(09), 3244A.

*Stanwyck, D. A., & Anson, C. A. (1989). *The Adopt-A-Student Evaluation Project: Final report.* Atlanta: Georgia State University, Department of Educational Foundations.

*Tierney, J. P., Grossman, J. B., & Resch, N. L. (1995). *Making a difference. An impact study of Big Brothers/Big Sisters.* Philadelphia: Public/Private Ventures.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

*Turner, S., & Scherman, A. (1996). Big brothers: Impact on little brothers' self-concepts and behaviors. *Adolescence*, *31,* 875–882.

*Valenzuela-Smith, M. (1984). The effectiveness of a tutoring program for junior high Latino students (Doctoral thesis, University of San Francisco, 1984). *Dissertation Abstracts International*, *46*(03), 0634A.

Wang, M. C., & Bushman, B. J. (1999). *Integrating results through meta-analytic review using SAS Software.* Cary, NC: SAS Institute.

*Watson, S. H. (1996). A study on the effects of four different types of mentoring plans using retired community individuals and college students as mentors on Hispanic "at risk" students' grades and attendance (Masters thesis, Stephen F. Austin State University, 1996). *Masters Abstracts International*, *35*(05), 1133.

Werner, E. E. (1995). Resilience in development. *Current Directions in Psychological Science*, *4,* 81–85.