Campbell Systematic Reviews
2013:10
First published: 02 September, 2013
Search executed: July, 2011

Mentoring Interventions to Affect Juvenile Delinquency and Associated Problems: A Systematic Review

Patrick Tolan, David Henry, Michael Schoeny, Arin Bass, Peter Lovegrove, Emily Nichols

THE CAMPBELL COLLABORATION

Title	Mentoring Interventions to Affect Juvenile Delinquency and Associated Problems: A Systematic Review
Authors	Patrick Tolan, David Henry, Michael Schoeny, Arin Bass, Peter Lovegrove, Emily Nichols Affiliations [1]University of Virginia
DOI	DOI: 10.4073/csr.2013.10
No. of pages	148
Citation	Tolan P, Henry D, Schoeny M, Bass A, Lovegrove P, Nichols E. Mentoring Interventions to Affect Juvenile Delinquency and Associated Problems: A Systematic Review Campbell Systematic Reviews 2013:10 DOI: 10.4073/csr.2013.10
ISSN	1891-1803
Copyright	
Contributions	Patrick Tolan (PT), David Henry (DH), and Michael Schoeny (MS) wrote the text of the protocol. Arin Bass (AB), Peter Lovegrove (PL), Emily Nichols (EM) and MS conducted the searches which PT directed. PT, PL and DH wrote the text of the review, DH, MS, and PT responded to editorial comments and those from external referees. We also want to acknowledge the generous aid in regard to databases and advice from Mark Lipsey, David DuBois, and Steven Aos.
Editors for this review	Editor: David B. Wilson Managing editor: Charlotte Gill
Support/funding	The review was conducted with support from the University of Illinois Institute for Juvenile Research, the University of Virginia's Youth-Nex Center and the Jerry Lee Foundation.
Potential conflicts of interest	None known.
Corresponding author	Patrick Tolan The U.Va Center to Promote Effective Youth Development University of Virginia 405 Emmet Street Charlottesville, VA 22903 USA Telephone: +1 434 243 9551 E-mail: pht6t@virginia.edu

Campbell Systematic Reviews

The Campbell Collaboration (C2) was founded on the principle that systematic reviews on the effects of interventions will inform and help improve policy and services. C2 offers editorial and methodological support to review authors throughout the process of producing a systematic review. A number of C2's editors, librarians, methodologists and external peer-reviewers contribute.

The Campbell Collaboration
P.O. Box 7004 St. Olavs plass
0130 Oslo, Norway
www.campbellcollaboration.org

Synopsis

Mentoring is one of the most commonly used interventions to prevent, divert, and remediate youth engaged in, or thought to be at risk for delinquent behavior, school failure, aggression, or other antisocial behavior. In this update we report on a meta-analytic review of selective and indicated mentoring interventions that have been evaluated for their effects on delinquency outcomes for youth (e.g., arrest or conviction as a delinquent, self-reported involvement) and key associated outcomes (aggression, drug use, academic functioning). Of 164 identified studies published between 1970 and 2011, 46 met criteria for inclusion. Mean effects sizes were significant and positive for delinquency and academic functioning with trends (marginal significance level) for aggression and drug use. Effect sizes were modest by Cohen's differentiation. However, there was heterogeneity in effect sizes across studies for each outcome. The obtained patterns of effects suggest mentoring may be valuable for those at-risk or already involved in delinquency and for associated outcomes. Comparison of study design (RCT vs. QE) did not show significant differences in effects. Moderator analysis showed larger effects when professional development was the motivation of the mentors for involvement, but not for basis of inclusion of participants (environmental vs. person basis of risk), presence of other interventions, or assessment of quality of fidelity. We also undertook the first systematic evaluation of key processes that seem to define how mentoring may aid youth (e.g. identification/modeling, teaching, emotional support, advocacy) to see if these related to effects. Based on studies we could code for the presence or absence of each as part of the program effort, analyses found stronger effects when emotional support and advocacy were emphasized. These results suggest mentoring is as effective for high-risk youth in relation to delinquency as many other preventive and treatment approaches and that emphasis on some theorized key processes may be more valuable than others. However, the collected set of studies is less informative than expected with quite limited specification about what comprised the mentoring program and implementation features. The juxtaposition of popular interest in mentoring and empirical evidence of benefits with the limited reporting of important features of the interventions is seen highlights the importance of more careful and extensive evaluations. Including features to understand testing of selection basis, program organization and features, implementation variations, and theorized processes for effects will greatly improve understanding of this intervention. All are essential to guide effective practice of this popular and very promising approach.

Abstract

1.1 BACKGROUND

Mentoring has drawn substantial interest from policymakers, intervention theorists, and those interested in identifying promising and useful evidence-based approaches to interventions for criminal justice and child welfare outcomes (Grossman & Tierney, 1998; Jekliek et al., 2002). Mentoring is one of the most commonly-used interventions to prevent, divert, and remediate youth engaged in, or thought to be at risk for, delinquent behavior, school failure, aggression, or other antisocial behavior (DuBois, Holloway, Valentine, & Cooper, 2002, DuBois, Portillo, Rhodes, Silverthorn, & Valentine, 2011). One account lists over 5000 organizations within the United States that use mentoring to promote youth wellbeing and reduce risk (MENTOR/National Mentoring Partnership, 2006). Definitions of mentoring vary, but there are common elements. For the purpose of this review, mentoring was defined by the following 4 characteristics: 1) interaction between two individuals over an extended period of time, 2) inequality of experience, knowledge, or power between the mentor and mentee (recipient), with the mentor possessing the greater share, 3) the mentee is in a position to imitate and benefit from the knowledge, skill, ability, or experience of the mentor, 4) the absence of the role inequality that typifies other helping relationships and is marked by professional training, certification, or predetermined status differences such as parent-child or teacher-student relationships. A total of 46 topic and methodologically eligible studies (out of 164 outcome reports) were identified for inclusion in the meta-analysis on delinquency and outcomes associated to delinquency: aggression, drug use, and academic achievement.

1.2 OBJECTIVES

This systematic review had the following objectives:

a) To statistically characterize the evidence to date on the effects of mentoring interventions (selective and indicated) for delinquency (e.g. arrest, reported delinquency), and related problems of aggression drug use, school failure.
b) To attempt to clarify the variation in effects of mentoring related to program organization and delivery, study methodology, and participant characteristics.
c) To help define mentoring in a more systematic fashion than has occurred to date to, in turn, help clarify how intervention processes suggested as compromising how mentoring has effects and other important considerations for future research..
d) To inform policy about the value of mentoring and the key features for utility.

1.3 SEARCH STRATEGY

This is an update of a review completed 4 years ago. In the original review search we benefitted from the authors of three meta-analyses on mentoring or related topics (1) DuBois et al. (2002) on mentoring in general, 2) Lipsey and Wilson (1998) on delinquency interventions in general, and 3) Aos et al. (2004) on interventions for delinquency and associated social problems) who provided databases on reports and coding approaches. In addition, we searched various databases including PsychINFO, Criminal Justice Abstracts, Criminal Justice Periodicals Index, Social Sciences Citation Index (SSCI), Science Citation Index (SCI), Applied Social Sciences Indexes and Abstracts (ASSIA), MEDLINE, Science Direct, Sociological Abstracts, Dissertation Abstracts, Database of Abstracts of Reviews of Effectiveness, and ERIC (Education Resources Information Center) and the Social, Psychological, Educational and Criminological Trials Register (SPECTR- in original search), the National Research Register (NRR, research in progress), and SIGLE (System for Information on Grey Literature in Europe). Finally, the reference lists of primary studies and reviews in studies identified from the search of electronic resources were scanned for any not-yet identified studies that were relevant to the systematic review. For this update we searched the same databases (except SPECTR as it no longer existed), surveyed pertinent journals and the reference lists of primary studies and reviews.

1.4 SELECTION CRITERIA

1. Studies that focused on youth who were at risk for juvenile delinquency or who were currently involved in delinquent behavior. Risk is defined as the presence of individual or ecological characteristics that increase the probability of delinquency in later adolescence or adulthood.

2. We included interventions focusing on prevention for those at-risk (selective interventions) and treatment (indicated interventions) that included mentoring as the intervention or one component of the intervention and at least measured impact of the program. We excluded studies in which the intervention was explicitly psychotherapeutic, behavior modification, or cognitive behavioral training and indicated provision of helping services as part of a professional role.
3. We required studies to measure at least one quantitative effect on one of the four outcomes (delinquency, aggression, substance use, academic achievement) in a comparison of mentoring to a control condition. Experimental and high quality quasi-experimental designs were included.
4. The review was limited to studies conducted within the United States or another predominately English-speaking country and reported in English and to studies reported between 1970 and 2011. We did not have resources for translating reports not reported in English.

1.5 DATA COLLECTION AND ANALYSIS

All eligible studies were coded using a protocol derived from three related prior meta-analyses, with 20% double-coded. The intervention effect for each outcome was standardized using well established methods to calculate an effect size with 95% confidence intervals for each of the four outcomes (if included in that study): delinquency, aggression, drug use and academic achievement. Meta-analyses were then conducted for each independent study within a given outcome (delinquency, aggression, drug use, and academic achievement). Effect sizes for each study were scaled so that a positive effect indicated a desirable outcome (i.e., lower delinquency, drug use, and aggression or higher academic achievement).

1.6 MAIN RESULTS

A total of 164 studies were identified as meeting inclusion criteria as focused on delinquency and mentoring. Of these, 46 met the additional criteria for inclusion in the quantitative analyses. 27 were randomized controlled trials and 19 were quasi-experimental studies involving non-random assignment, but with matched comparison groups as was described above. Twenty-five studies reported delinquency outcomes, 25 reported academic achievement outcomes, 6 reported drug use outcomes, and 7 reported aggression outcomes.

Main effects sizes were positive and statistically significant for all four outcomes. Some studies showed effects that were not significant and a few reported negative effects. For each outcome there was substantial variation in effect size, too. Average effects were larger for delinquency than for other outcomes. When moderation was tested, there was considerable variation in effect sizes of studies that were similar in regard to the presence of a given moderating influence.

We compared effect sizes of those studies that were random assignment experimental designs with those that were quasi-experimental using meta-regression and found no evidence of differences in effect sizes. We conducted moderator analyses to determine whether effects found differed by 1) criteria for selecting participants, 2) presence of other components along with the mentoring intervention, 3) motivation of mentors for participation, or 4) assessment of quality or fidelity of implementation of the intervention. We also conducted moderator analyses to test for outcome differences by the presence or absence of four theorized key components of mentoring interventions. The relatively limited information about potential moderating characteristics extractable from many reports and the limited number of reports with extractable information led us to combine effects across all four outcomes to enable adequate power and in combination to our directional expectations for moderators to test significance using a one-tailed test ($p < .05$). For these analyses, we averaged effect sizes within a given study if more than one outcome of interest was reported. We also conducted analyses to check for bias in effects due to type of outcome, and found no suggestion of bias.

We found evidence for moderation when professional development was a motive for becoming a mentor. There was also moderation of the effect size when mentoring programs emphasized either of two theorized components: emotional support or advocacy. Effect sizes did not differ by whether or not the program emphasized the other two key components: modeling/identification or teaching, nor by whether other components were used, how risk was defined (environmental versus individual characteristics) or if fidelity/adherence of implementation features were assessed.

1.7 REVIEWERS' CONCLUSIONS

This analysis of 46 studies on four outcomes measuring delinquency or closely related outcomes of aggression, drug use, and academic functioning suggests mentoring for high-risk youth has a modest positive effect for delinquency and academic functioning, with trends suggesting similar benefits for aggression and drug use. Effect sizes varied more for delinquency and academic achievement than for aggression and drug use. We did not find a significant difference in effect size by study design (RA vs. QE) or by whether or not fidelity was assessed. We identified some characteristics that moderated effects that provide additional understanding for further studies and program design. Effects tended to be stronger when professional development was an explicit motive for participation of the mentors. Of four processes theorized as comprising the methods of effects in mentoring, we found evidence for significantly larger effects when emotional support and advocacy were emphasized. Although these findings support viewing mentoring as a useful approach for intervention to lessen delinquency risk or involvement, limited description of content of mentoring programs and substantial variation in what is included as part of mentoring efforts detracts from better understanding about what might account for the benefits. The valuable features and most promising approaches cannot be ascertained with any certainty. In fact, the body of studies is remarkably lacking in description of key features, program design organization, and theorized processes of impact that are typically provided in empirical reports of intervention effects. Our judgment is also that there does not seem to be much progression in quality of details in reports over the time period studied here. Given the popularity of this approach, the promise of benefits should be seen as a strong argument for a concerted effort through quality randomized trials to specify the theoretical and practical components for effective mentoring with high-risk youth. Concordantly, lacking such features, further trials may not add useful knowledge.

1 Background

Mentoring is one of the most commonly-used interventions to prevent, divert, and remediate youth engaged in, or thought to be at risk for, delinquent behavior, school failure, aggression, or other antisocial behavior (DuBois, Holloway, Valentine, & Cooper, 2002). It is the centerpiece of the work of the Boys and Girls Clubs of America. A recent account lists over 5000 organizations within the United States that use mentoring to promote youth wellbeing and reduce risk (MENTOR/National Mentoring Partnership, 2006). In fiscal year 2011 it is estimated approximately $100 million in federal support and research funds were dedicated to youth mentoring (DuBois et al., 2011).

Definitions of mentoring vary, but there are common elements that can be identified across definitions (DuBois & Karcher, 2005, DuBois, et al., 2011). Most commonly the central feature is a one-on-one relationship between a provider (mentor) and a recipient (mentee) for the potential of benefit for the mentee. For the purpose of this review, mentoring will be defined by the following 4 characteristics: 1) interaction between two individuals over an extended period of time, 2) inequality of experience, knowledge, or power between the mentor and mentee (recipient), with the mentor possessing the greater share, 3) the mentee is in a position to imitate and benefit from the knowledge, skill, ability, or experience of the mentor, 4) absence of role inequality between provider and recipient that typifies most helping or intervention relationships whether based in professional training or certification of the provider or as occurs inherent in parent-child, teacher-student, or other professional-client relationships. Thus, mentoring differs from professional-client relationships such as counseling or therapy, and from parenting or formal educational relationships.

When applied to delinquency and other similar outcomes, mentoring usually involves older, usually adult, persons in the community who provide opportunities for imitation, gaining advice, pleasurable recreational activities that show care and interest in the mentee, and emotional support, information, and advocacy through a one-to-one relationship. Such opportunities are thought to foster healthy development and diversion from risk-elevating activities and attitudes (Jekielek, Moore, Hair, & Scarupa, 2002; Rhodes, Spencer, Keller, Lian, & Noam, 2006).

Over the past twenty years, mentoring has drawn substantial interest from policymakers, intervention theorists, and those interested in identifying promising and useful evidence-based approaches to interventions for criminal justice and child welfare outcomes (Grossman & Tierney, 1998; Jekielek et al., 2002). This has included a substantial investment in the United States and elsewhere in support for implementation of mentoring, a professional organization dedicated to advancing quality of and use of mentoring, and a proliferation of mentoring programs (MENTOR/National Mentoring Partnership, 2006). The popularity and extensive anecdotal praise for mentoring makes it important to have sound, evidence-based, understanding of its promise. While prior meta-analyses and moderation tests of specific interventions can point to some potentially important features, none of these analyses have focused on mentoring as an intervention for youth at risk for delinquency. In this study, we conduct a meta-analytic review of mentoring interventions that have been evaluated for their effects on delinquency (e.g., arrest or conviction as a delinquent, self-reported involvement) and three outcomes (aggression, drug use, academic achievement) that often co-occur with delinquency, share risk factors, are often also targets of delinquency interventions and show effects from such efforts (Tolan, 2002).

Unlike many types of intervention, there are a substantial number of studies that evaluate the effects of some form of youth mentoring (DuBois et al., 2011; Rhodes, Bogat, Roffman, Edelmena, & Galasso, 2002). Critical reviews have focused on the potential benefits of mentoring and characteristics that might be associated with positive effects from mentoring (Hall, 2003; Rhodes, 2002). More recently, several meta-analyses have considered mentoring programs in relation to youth risk, including delinquency (Aos, Lieb, Mayfield, Miller, & Pennucci, 2004, DuBois et al., 2002; DuBois et al., 2011; Lipsey & Wilson, 1998). Thus, unlike some areas of intervention for delinquency and related problems, the accumulated literature on mentoring is substantial and has had conceptual and statistical scrutiny. None of the meta-analyses to date correspond exactly with the focus of the present review, but they were very helpful in planning this review. They suggest standards against which to evaluate the completeness of study inclusion and choices about coding and methodological requirements.

Many of the conceptual reviews have been focused on the potential of mentoring as a general approach to youth development promotion and to reduce risk among high-risk populations (Jekielek et al., 2002; Rhodes, 2002). The meta-analysis by DuBois et al. (2002) focused on mentoring efforts related to youth development. Although there was differentiation of "problem-behavior" from other outcomes (e.g. educational attainment, vocational) there was not clear emphasis on delinquency indicators as a separate area. The follow-up/updating to that review in 2011 (DuBois et al., 2011) utilized a similar general outcome category (conduct problems). Notably, DuBois et al (2011) report that effect sizes were larger for programs serving youth involved in problem-behavior than for those with other bases for inclusion

and also for those with higher levels of environmental or individual risk. Lipsey and Wilson (1998) organized their review around an interest in serious juvenile offenders. Therefore, inclusion was not about delinquency risk in general, precursors such as aggression level, or related outcomes such as substance use or academic functioning. Also, the interventions were coded in such a way that interventions that included mentoring in an array of interventions could not be distinguished from those that focused primarily or exclusively on mentoring. Mentoring was denoted by its mention in the description of a study, but often was considered as one member of a class of interventions with similar features. Aos et al. (2004) undertook their meta-analysis to inform a state legislature about the potential impact, costs and benefits of many empirically tested interventions for delinquency and other youth problems such as early pregnancy. Thus, their emphasis was on specific programs rather than mentoring as a general approach. Moreover, that review only examined the relative effect sizes in relation to costs and potential cost savings rather than the usual focus on methodological issues and other moderators of effects. In addition, they were interested in programs with a high level of empirical support for effects, so that their inclusion criteria were more restrictive than was used here.

The aforementioned conceptual and statistical reviews provided excellent perspectives on mentoring evaluation and valuable benchmarks for guiding this review. In addition, they provided strong data bases from which to organize this review. Because of the generous sharing of information about content and methods by these reviewers (including access to their databases in some cases), this review was able to build efficiently from their prior efforts in determining coding. These prior reviews also helped reduce worry about file drawer and grey material that might be important to consider but not found without thorough searching. Of course, we conducted an independent search to verify the applicable literature, published or not.

The accumulated reviews and the variations in the studies they included also point to the value of this review. Each suggested mentoring programs could have important effects on delinquency and related outcomes. In the DuBois et al. review (2002), the overall category of problem behavior, which includes delinquency, had the largest effect sizes of any outcome category. This was confirmed in the updating of that review in the sense that average effect was equal to or close to each of the other categories considered (mean *d* =.21; DuBois et al., 2011). However, that review focused on mentoring in general and effects irrespective of sample selection basis. That review and others noted the variation in effects even among well-designed and completed studies; variation that undercut confidence in the mean effect findings. In fact, the field is marked by mixed results (significant positive and negative effects) among the methodologically stronger studies (e.g. McCord, 1992; O'Donnell, Lydgate, & Fo, 1979). As the DuBois et al (2002, 2011) reviews excluded the McCord Cambridge-Somerville and another major mentoring study, the

Diversion Project of Davidson and colleagues (Davidson & Redner, 1988), the implications for mentoring as a criminal justice intervention are not clear. Both are studies that carried likely substantial impact on overall effect estimates and for design impact or moderator interpretation. Also, a scan of the literature at the outset of the review showed several new pertinent studies since the prior reviews. The present study focuses on mentoring as focused on youth at-risk for delinquency, a more specific population focus than prior reviews of mentoring. Finally, with this updating this review includes studies not considered by the other mentoring reviews.

Understanding Mentoring Effects

While a relatively large number of studies with some minimal evaluation design features have been found and utilized in prior evaluations of mentoring, there are characteristics of this field that have limited how informative reviews and meta-analyses have been. For example, the most often considered intervention feature is the extent of matching on similarity of demographic characteristics and interests of the mentor and recipient and that the quality of the personal relationship not be marked by dissatisfaction (referred to as mentor relationship quality; see Rhodes, 2005; DuBois et al., 2011). Yet, these are merely post-hoc identified variables differentiating effect sizes. They do not directly or indirectly indicate what are the processes through which mentoring has its effects or suggest what it is about mentoring that might make it different from many other helping relationships (MENTOR/National Mentoring Partnership, 2009). A major limitation of the field, and perhaps of progress in understanding more about how and why mentoring shows positive effects is the lack of specificity in describing the activities comprising a given mentoring intervention and perhaps more importantly tying activities and practices to theorized key processes through which positive impact is thought to occur. This limits ability to compare program features that may explain variations in effects as well as limiting ability to tie programs to theories about interpersonal processes that could explain how mentoring has impact (DuBois et al., 2011; Rhodes, Grossman, & Resch, 2000). Among the reports and reviews there is considerable variation in what activities are considered mentoring essentials and which are optional (DuBois & Karcher, 2005; (MENTOR/National Mentoring Partnership, 2009). At the same time there is not much attention to and little certainty about what constitutes a mentoring intervention and might distinguish mentoring from other helping relationships (Rhodes et al., 2000). In addition, more understanding of key processes of a given program could improve ability to compare across programs and cumulatively point toward the important or necessary components that define a program as mentoring (Roberts, Liabo, Lucas, DuBois, & Sheldon, 2004).

Limited intervention description may be because mentoring arose as a voluntary and "indigenous" approach to youth intervention. Thus, many mentoring efforts arose

within a given setting without intention to formalize and standardize performance and activities. The practitioners who developed their particular approach may have had less training in and interest in formal evaluation features. As a field of intervention services and as a research focus, there appears to be mixed interest in facilitating more formal operations that will yield more informative and comparable results (MENTOR/National Mentoring Partnership, 2006). Also, because one common basis for mentoring is a view that the positive influence of an interested person providing a supportive relationship is what is helping, there may well be less interest in trying to specify what activities and processes constitute mentoring and what among these could explain any benefits derived. For all of these reasons formalized protocols and systematic training approaches may not have been a priority. Consequently the body of research is remarkable in the limited emphasis on systematic description of intervention content, description of intended processes through which effects are expected, and in important features of implementation and providers. There seems to remain limited valuing of and perhaps even some reluctance to aim for continuity across the field or specificity in applying and describing mentoring efforts that might facilitate scientific understanding of effects. Hence, there are few training, implementation, and dosage parameters that can be identified as having consensus. There are few indications of what is considered essential or critical for mentoring and helpful in distinguishing mentoring from other helping relationships and approaches. Similarly, the reports reviewed here continue an unfortunate tradition of having limited information by intervention science standards and are less informative than needed about what may account for benefits accrued. Overall, greater interest in relating to a common set of principles, theorized processes, or requisite structures and components would seem an advance that could serve the field well. Thus, one of the goals of this effort and one that is a different emphasis than prior meta-analyses was to code, to the extent possible, comprising activities, mentor selection/motivation, and training or implementation features as viewed from an intervention research lens. A second was to theorize four processes that are often mentioned or pointed to as how mentoring affects youth positively and that in total can distinguish mentoring from other forms of helping interventions. Of interest was that this combination do not as a group also from the critical processes of other helping interventions such as teaching or psychotherapy. Emphases on these two interests (intervention features and key processes) could potentially help advance understanding of mentoring, effects found, and potential for further study and use.

Important Intervention Features Affecting Mentoring Impact

One aspect of mentoring intervention characteristics given substantial attention is the implication that a strong personal relationship between the mentor and mentee is a key to any benefits derived (DuBois et al., 2002; Rhodes et al., 2006). Thus one advance in the field is to assess how positive and engaging the relationship is between the mentor and mentee (Rhodes et al., 2006). For example, DuBois et al. (2011) report larger effect sizes when matching of mentor and youth was based on

shared interests; presumably this improves the likelihood of a good relationship. In most cases, a corollary is that the mentor is undertaking this activity, not as a professional in the helping or social service professions, but because of personal interest or sense of duty, often as a volunteer (Rhodes, 2002). When a person with professional background or duties to provide such services offers mentoring, the emphasis is more on the relationship and the personal interest in the mentee than on specific skills, activities, or formal protocols. Thus, it has been noted that one limitation of mentoring may be that providers may be less accountable as they are volunteers and/or may not be well prepared for challenges of developing and maintaining a relationship with sometimes challenging and less appreciative youth (Grossman & Tierney, 1998). In contrast, it may be that motivation that is not personal, that is for professional advancement or as a paid position might be expected to lessen the personal commitment and connection thought to spark effective mentoring. More understanding of how different reasons for undertaking mentoring influence effects would help with understanding effect variations and provide direction for improving impact. We test for differences by motivation of mentor for engaging in this work.

A second area of some importance in understanding how to direct mentoring efforts is the effect of structuring of the effort and expecting fidelity to an approach. While it is increasingly recognized that training in skills and expectations are important for mentoring, there is much less clarity about what is important to expect. Mentoring has been characterized as growing out of a mentor's commitment to youth (Rhodes, 2002) with the accompanying implication that structuring the activities and processes to be ensured would detract from the individualistic authentic engagement that carries the benefits. In contrast, research on other forms of intervention have not supported such a view, pointing to more clear expectations and fidelity prescriptions as promoting larger effects (Tolan & Gorman-Smith, 2003). Thus, the extent to which there is emphasis on following these procedures and principles thought to be helpful should relate to effect levels. Therefore in this review we examine if assessment of fidelity relates to effect size.

A third question of importance about mentoring is the relative value of mentoring as a high-risk selective and/or indicated approach rather than as a universal intervention (Tolan & Guerra, 1994). Mentoring studies have been applied to high-risk, identified, and general populations of youth. There is some indication the effects might be greater for higher risk youth, although the results are not fully consistent (DuBois et al., 2011). Some have argued that mentoring represents an alternative view about youth risk, a focus on promoting healthy or positive development through strengthening abilities rather than minimizing exposure to risk or remediation of undesirable behavior and characteristics (Jekielek, et al, 2002). Also, it may be that mentoring that is developed for and applied to high-risk youth has impact for that population that programs for non-delinquent or general population youth do not. There is evidence that preventive effects for high-risk

youth may be quite different from those accrued for the general population (Tolan & Gorman-Smith, 2003). For example, it may be that mentoring is not valuable in affecting delinquency or related outcomes of high risk youth because it is not structured enough and focused on multiple risk factors thought to drive that behavior (Lipsey & Wilson, 1998). Thus, there is a policy interest in whether targeting high-risk youth (selective inclusion) is useful. Therefore, the review undertaken here was focused on youth defined as high-risk for or already engaged in delinquency (Tolan & Gorman-Smith, 2003).

Similarly, as others have noted it is common for mentoring to occur as part of a multi-component program, whether as one of several components or as a central focus augmented by additional supporting activities (Aos et al., 2004). This leaves open an important question of the extent to which effects attributed to mentoring might actually be coincidental inclusion with other effective components. It also leaves undifferentiated to what extent it matters if the delivery with other components is simply as one of a set of substantial program features or if the program is primarily mentoring with some augmentation to help support and enhance the mentoring impact. These questions of interest suggest coding of these features, where discernible, might improve understanding of the value of mentoring.

Identifying Potential Key Processes Defining and Differentiating Mentoring

In addition to these features of intervention organization that have been of interest in characterizing mentoring as a field of intervention and in furthering the evaluation knowledge about mentoring, there is an important but almost unattended to issue of what processes are typical of and constitute mentoring as an intervention. Are there activities or underlying purposes of activities that are common to mentoring or that might vary across mentoring programs and in doing so help account for differences in effects? As noted above, theoretical summaries of the field and attempts to relate mentoring to prevention science, developmental psychopathology, and/or youth development literature in general have suggested some likely key features of mentoring (Lipsey & Wilson, 1998; McCord, 1992; Tolan & Guerra, 1994). These processes are differentiated from the attention to the mentor-mentee relationship that has dominated evaluation of mentoring (Rhodes, 2005, DuBois & Karcher, 2005). The latter represents an aspect of connection that while important as a basis for mentoring is a common basis for any influence relationship.

Through systematic review of theoretical organization of process models of mentoring (e.g., (MENTOR/National Mentoring Partnership, 2009; Rhodes, 2005), indices utilized by DuBois et al. 2002 in examining how effect size of mentoring related to score on best practices index, components described in programs with significant effects (e.g., Davidson & Redner, 1988), and qualitative analyses of

mentoring relationships (e.g. Deutsch & Spencer, 2009) we organized a set of processes that seemed to occur across mentoring programs, whether explicitly described or implicit in the activities utilized. In addition, we compared mentoring to other helping interventions to identify distinguishing features. For example, mentoring is distinguished from psychotherapy by the non-professional relationship and the lack of emphasis on mental health problem alleviating. From these multiple bases we identified four processes as central to mentoring: 1) identification of the recipient with the mentor that helps with motivation, behavior, and bonding to conventions; 2) provision of information or teaching that might aid the recipient in managing social, educational, legal, family, and peer challenges; 3) advocacy for the recipient in various systems and settings; and 4) emotional support and friendliness to promote self-efficacy, confidence, and sense of mattering (DuBois et al., 2002; DuBois et al., 2011; Rhodes et al., 2002). These processes are frequently mentioned individually as potential bases for mentoring benefits. More recently some attention has been given to how advocacy within mentoring can affect impact. DuBois et al. (2011) report when advocacy was considered a mentor function effect sizes averaged .07 standard deviation units larger than when not. Also, several of the more fully described efforts point to one or more of these processes as intended elements of the mentoring. Understanding of whether emphasis on such processes relates to effects is one intended contribution of this review. Therefore, we coded studies for evidence of each key process driving or comprising the intervention to permit examining how their inclusion may have affected outcome.

Prior Evaluation of Features Affecting Mentoring Impact

DuBois et al. (2002, 2011) recognized many of the issues related to advancing and deepening understanding of mentoring effects and incorporated coding of several of these features into their meta-analysis. In DuBois et al. (2002), they denoted an index of what could be considered best practices in youth mentoring based on recommendations of prior reviews and recommendations for establishing effective mentoring programs, such as the National Mentoring Working Group (1991) and coded to the extent possible from source data, each intervention report (DuBois et al., 2002). They included 11 program features to mark how methodic inclusion in the program was, whether mentors and mentees are matched on demographic characteristics, how structured or prescribed activities were, and the frequency or extent of contact. These codes were then amalgamated into an overall index of extent of desired features. While this represents an informative advance about how the extent of features considered useful for good mentoring related to effect size, because it is a single score across many areas it cannot indicate the importance of specific features. Also, it may have obscured how many of the reports did not have adequate reporting to fully assess the 11 features.

We attempt to build on efforts of DuBois et al. (2002) to code theoretically and empirically linked valued characteristics, activities, and organization by focusing on

the moderating effects of each of several key features related to 1) selectivity in inclusion (high risk versus universal or no selectivity within the population); 2) explicit attention to presence of four key processes such as modeling, emotional support, advocacy, and teaching; 3) whether or not mentoring is a stand-alone approach in that study or was undertaken along with some other components: 4) the motivation of the mentors in participating; and 5) the extent to which quality of work and fidelity were assessed or emphasized. This coding was considered useful for suggesting what might differentiate mentoring from other similarly intended youth interventions. Despite prior identification of specificity of such features as a major limitation of the mentoring literature (Tolan & Guerra, 1996), we did not find much improvement over time in the ability to determine details needed to code many of these features for this review. We had to limit our analyses to those features that could be coded for enough studies to enable some useful comparison.

2 Objectives of the Review

This updating of a prior systematic review had the following objectives:

1. To statistically characterize the evidence to date on the effects of mentoring interventions (preventive and treatment) for delinquency (e.g. official records and self-reported), and the associated problems of aggression drug use, and school failure.
2. To examine the heterogeneity of effects for each outcome and the role of design (RA vs. QE) in the effects found.
3. To examine the relation of a few key aspects of mentoring interventions (e.g. selection vs. universal inclusion, mentor motivation, quality and fidelity control, presence of important features of mentoring, and presence of other interventions) to effects found.
4. To suggest important features of existing literature to be further developed and supported to improve how informative evaluations can be and to increase comparability across mentoring efforts.
5. To identify gaps in this research area and make recommendations for further research.
6. To inform policy about the value of mentoring and the key features for utility.

3 Methods

In order to provide a review that is as free of bias as possible, we adopted a systematic review strategy for the research on the effects of mentoring interventions as guided by Campbell Collaboration standards and employed in the original review. This report is an update of the prior Campbell Review that covered reports between 1970 and 2005 (www.campbellcollaboration.org/lib/download/238/). This updates the review for reports available through July of 2011.

Search strategy for identifying relevant studies

Three authors have conducted prior meta-analyses on mentoring or related topics: 1) DuBois et al. (2002) on mentoring in general, 2) Lipsey and Wilson (1998) on delinquency interventions in general, and 3) Aos et al. (2004) on interventions for delinquency and associated social problems. Prior to conducting this review, each of these authors allowed us access to some of the materials used in their analyses. Drs. Lipsey and Aos and their colleagues released the actual databases used for their analysis. We found that one or more of these authors had already located many of the studies to be included in this analysis. However, we conducted our own review to locate studies done since these earlier reviews were completed and to locate other studies, including those that were unpublished at the time of these previous analyses. During the search phase, abstracts were reviewed and studies that did not include the target outcomes or were clearly not of experimental/quasi-experimental design were excluded from further consideration. Full-text copies of the remaining 164 studies were then obtained. We used dates, sample sizes, authorship, and information provided on studies to determine whether two effects on the same outcome came from the same study. We did not count effect sizes at different follow-up points as independent effects, using the effect most close to post-test for these analyses.

Search terms and databases

We based our search terms on those used by prior meta-analyses. We used a combination of terms in searching electronic databases and research registers. Table 1 shows the search terms used, although slight deviations in key words (including derivative forms of the listed terms) required modification to achieve equivalent searches in some databases (e.g., choosing a broader search term when a narrower term was not supported in the database). We also provide details of combinations of

the search terms and some examples of resulting search combinations (shown in the inner cells) in Table 2. We searched the databases using combinations of terms, each of which contained: 1) one of four outcomes (and derivative forms of these terms): delinquency, aggression, substance use, or academic achievement; 2) a cognate of mentoring; and 3) a cognate of intervention

Databases searched

Databases were selected based on their potential relevance to the topic and to the outcomes of delinquency, academic achievement, aggression, and substance use more generally. The databases searched included PsychINFO, Criminal Justice Abstracts, Criminal Justice Periodicals Index, Social Sciences Citation Index (SSCI), Science Citation Index (SCI), Applied Social Sciences Indexes and Abstracts (ASSIA), MEDLINE, Science Direct, Sociological Abstracts, Dissertation Abstracts, Database of Abstracts of Reviews of Effectiveness, and ERIC (Education Resources Information Center). The following research registers were also searched: the Social, Psychological, Educational and Criminological Trials Register (SPECTR (for original search, not used in update), the National Research Register (NRR, research in progress), and SIGLE (System for Information on Grey Literature in Europe). Finally, the reference lists of primary studies and reviews in studies identified from the search of electronic resources were scanned for any not yet identified that were relevant to the systematic review. All searches covered until July 2011.

3.1 CRITERIA FOR INCLUSION AND EXCLUSION OF STUDIES IN THE REVIEW

Only studies that satisfy all of the following inclusion criteria and none of the following exclusion criteria were included in this review:

Outcomes measured

We focused this systematic review on outcomes related to juvenile delinquency. We included studies with outcome measures of juvenile delinquency, reported by the individual or by others, or derived from archival sources such as arrest or juvenile court records. We also included studies focusing on precursors of delinquency such as aggression or high levels of externalizing problems and studies with two outcomes that are correlated with and frequently co-occur with criminal involvement or delinquency risk (drug abuse and academic achievement/ school failure). As noted above, the specific terms for each outcome are provided in Table 1.

Types of participants

Juvenile delinquency is typically defined as antisocial or criminal behavior by persons under age 18 (Tolan, 2002). In this systematic review of mentoring interventions, we included studies that involved youth who were included because they were currently showing behavior that would constitute juvenile delinquency or

were identified and included because they were "at-risk" for juvenile delinquency. At-risk is defined as the presence of individual or ecological characteristics that increase the probability of delinquency in later adolescence or adulthood (Tolan, 2000). Ecological characteristics include family and parenting influences on behavior, residence in neighborhoods with high levels of poverty or crime, exposure to gangs, and other social setting factors (Tolan & Gorman-Smith, 2003). Individual characteristics include high scores on screening measures for aggression, evidence of oppositional defiant or conduct disorders, school failure, or attitudes and beliefs consistent with elevated use of aggression or antisocial behavior (Farrington, 2004). Demographic characteristics were not considered as designating at-risk for consideration for inclusion here. Thus, a study that targeted a demographic group even if doing so because they are considered at risk was not included unless selection met our criteria otherwise.

Intervention Type

We included interventions focusing on prevention and treatment (referred to as selective and indicated population interventions). In the initial phase of study selection, we sought out any studies that described their interventions as mentoring, that mentioned mentoring as any part of their intervention strategy, or had interventions characterized by any of the four characteristics noted above, whether or not they specifically mentioned mentoring.

Regarding the defining characteristic of absence of formalized role inequality, previous reviews have differed on the inclusion of studies using professionals as mentors. DuBois et al. (2002) excluded interventions using professional providers, with the exception that some studies that employed mental health professionals as mentors were included under certain conditions (see DuBois et al., 2002; Rhodes, 2002 for those criteria). This appears to also have been the approach used in the updated meta-analysis by DuBois et al., 2011. We differed from these prior reviews by including studies with mental health providers as mentors if their involvement was unstructured or limited to a non-specific or support intervention (not psychotherapeutic). Functionally this means inclusion here of some critical studies for the current focus that were not included in the DuBois review, such as the McCord Cambridge-Somerville study (McCord, 1978, 1979).

We then excluded studies in which the intervention was explicitly psychotherapeutic, behavior modification, or cognitive behavioral training. Although we included studies in which mentoring was done as a part of another structured intervention, those studies that were conducted without providing results for the mentoring intervention separately were coded as including either an additional primary intervention (i.e., a major component in addition to mentoring) or an additional secondary intervention (i.e., a minor component in addition to mentoring).

In addition to requiring that studies investigate the effects of a mentoring intervention, as described above, we followed three additional criteria based on those used by Lipsey and Wilson (1998) in their meta-analysis of intervention effects on delinquency. We only included studies that measured at least one quantified outcome variable for the outcome of interest among the four considered here and that provided sufficient data to allow calculate an effect size and decipher its direction. When studies measured a delinquency-related outcome but did not report sufficient detail to allow calculation of an effect size, we attempted to contact the author to obtain additional information. Because of access to the Aos and Lipsey databases we had a relatively complete rendering of the studies from which such information could be extracted. There were, therefore, very few studies that we were uncertain about whether additional information was obtainable.

Research Design

The second criterion for inclusion in this review was that the study design involves a comparison that contrasted an intervention condition involving mentoring with a control condition. Control conditions could be "no treatment," "waiting list," "treatment as usual," or "placebo treatment". To ensure comparability across studies we made an a priori rule to not include comparisons to another experimental or actively applied intervention beyond treatment as usual. However, there were no such cases among the studies otherwise meeting criteria for inclusion.

We coded studies according to whether they were experimental or quasi-experimental designs. To qualify as experimental or quasi-experimental for the purposes of this review, we required each study to meet *at least one* of three criteria: 1) random assignment of subjects to treatment and control conditions or assignment by a procedure plausibly equivalent to randomization; 2) Individual subjects in the treatment and control conditions were prospectively matched on pretest variables and/or other relevant personal and demographic characteristics; 3) Use of a comparison group with demonstrated retrospective pretest equivalence on the outcome variables and demographic characteristics as described below

Randomized controlled trials that met the above conditions were clearly eligible for inclusion in the review. At the other end of inclusion eligibility, single-group pretest-posttest designs (studies in which the effects of treatment are examined by comparing measures taken before treatment with measures taken after treatment on a single subject sample) were never eligible. A few nonequivalent comparison group designs (studies in which treatment and control groups were compared even though the research subjects were not randomly assigned to those groups) were included. Such studies were only included if they matched treatment and control groups prior to treatment on at least one recognized risk variable for delinquency, had pretest measures for outcomes on which the treatment and control groups were compared and found to be essentially equivalent. We required that non-randomized quasi-experimental studies employed pre-treatment measures of delinquent, criminal, or

antisocial behavior, or significant risk factors for such behavior, that were reported in a form that permitted assessment of the initial equivalence of the treatment and control groups on those variables.

Time Period and English Language Criteria

We limited the review to those studies conducted within the United States or another predominately English-speaking country and reported in English. This was because we did not have resources for translation of studies not published in English and the vast majority of programs were conducted in the United States. Juvenile subjects did not need to speak English. A study conducted in the United States or Canada with resident Hispanic youth, for example, could have been included.

We limited the review to studies published since 1970. The time frame between 1970 and the present (time of completion of search to conduct coding, 2011) is consistent with start of the time interval used by the review of the literature on delinquency conducted by Lipsey and Wilson (1998) and others. This also is the time period for most almost all the available studies with the necessary information and design features to be included in this review.

Coding of Article Characteristics

We double-coded 20% of the new articles (N=32), and calculated inter-coder reliability coefficients for study type (e.g., randomized trial), study quality, participant selection criteria (e.g., individual or behavioral risk), mentor motivations (e.g., survivor of abuse, professional development), and intervention components (e.g., modeling, teaching) using Cohen's *kappa*. We found high reliabilities for study type ($\kappa = 1.0$), study quality ($\kappa = .93$), and selection criteria ($\kappa = .81$). Coders easily determined some mentor motivations such as personal experience that connected to the youth needs (e.g. experienced abuse) ($\kappa = .90$), but were less certain with topics such as civic duty or professional development ($\kappa = .68$). Not all categories were coded in the random sample of studies that were double coded. For example, of the mentoring components (modeling/identification, teaching, and emotional support) only modeling was found in the studies randomly selected for double coding. Final kappa reliabilities all were above .6, a level Landis and Koch (1977) suggested represented full agreement. Coders sought consensus with their supervisors, particularly on difficult-to-code categories such as mentor motivations. If this could not resolve differences then author Schoeny made a decision about categorical coding.

Effect sizes for outcomes were also double-coded for 20% of the new articles. There were no substantial variations in these ($r = .99$) with only one disagreement. As with other coding decisions we first attempted to resolve bases for differences (e.g. technical inconsistency that if corrected removed difference). We had a protocol in place to then structure discussion of differences to attempt to reach consensus. If necessary a decision would then be made by either the first author (Tolan) or author

Henry. Given the level of agreement we did not have to proceed past the technical comparison and a brief discussion to reach consensus.

We conducted a separate meta-analysis for each outcome (delinquency, aggression, drug use, academic achievement). Each grouping of studies was based on the outcome, such that some studies might be included in more than one meta-analysis due to measuring more than one outcome. Thirteen studies reported more than one outcome, four of which had three outcomes. A single outcome measure was used for each study for a given outcome category. No studies reported multiple measures of a single outcome (e.g., multiple measures of delinquency or aggression).

Statistical Procedures.

Effect Size Calculations: For this study we used inverse-variance meta-analysis with a random-effects model, performed and plotted through the *metagen* package in the *R* statistical language. The random effects model addresses the research question of whether the average effects of an intervention in the population are significantly different from zero (Bailey, 1987; Raudenbush, 1994).

The inverse variance method, as its name suggests, weights individual studies by the inverse of variance of their effect size. Thus, this method requires the calculation of standard errors of the effect sizes. For this purpose, we estimated variances for each effect size according to Hedges and Olkin's (1985 , p. 86) Formula 14:

$$\hat{\sigma}_{d_i}{}^2 = \frac{(n_e + n_c)}{(n_e * n_c)} + \frac{d_i{}^2}{(n_e + n_c)}$$

Where σ_{di}^2 is the estimated variance of the effect size, n_e is the number of experimental subjects, n_c is the number of control subjects, and d_i^2 is the square of the effect size of the study.

The standardized mean difference effect sizes of the interventions under evaluation were calculated in units of Hedges' (1981) g. For studies reporting means, standard deviations, and *N*s of numeric data, the effect size was calculated by dividing the treatment difference less the control difference over the pooled treatment and control standard deviation:

$$SMD = \frac{((ME2 - ME1) - (MC2 - MC1))}{\sqrt{\frac{\left((NE2 - 1)SE1^2 + (NC2 - 1)SC1^2\right)}{(NE1 + NC1 - 2)}}}$$

,

where: M = mean S = standard deviation
E = treatment C = control
1 = pretest 2 = posttest

For studies that reported dichotomous outcomes, we calculated odds ratios and converted them into an equivalent standardized mean difference effect size estimate (Lipsey & Wilson, 1998). Chinn (2000) noted that dividing the natural log of an odds ratio by $\pi/\sqrt{3}$ produces an excellent approximation of the standardized mean difference effect size.

We also applied a correction to all effect sizes that compensates for small sample bias:

$$g^* \approx (1 - \frac{3}{4(n_1 + n_2) - 9})g$$

We examined funnel plots from each meta analysis for visual evidence of asymmetry, and conducted Egger tests (Egger, Smith, Schneider, & Minder 1997) to obtain a statistical test for asymmetry. The Egger test fits a regression of the normalized effect estimate (estimate divided by its standard error) against precision (the reciprocal of the standard error of the estimate).

We conducted analyses to determine whether the effects of the mentoring interventions varied by five key aspects of the intervention approach and characteristics. Potential moderators that were tested were:

1) selectivity in inclusion (high individual risk, high environmental risk, or no such selectivity)
2) whether or not mentoring is a stand-alone approach in that study or was undertaken along with a) some other major intervention components or b) some relatively minor add-ons
3) the motivation of the mentors in participating (civic duty, professional development, own experience)
4) the extent to which quality of work and fidelity were assessed or emphasized.
5) explicit attention to presence of four key processes: modeling/identification promotion, emotional support, advocacy, and teaching

Inspection of the coding across studies indicated that we had to simplify some moderation analyses due to sparse or no studies noting a particular characteristic of interest. For selection of participants, none of the interventions were coded as a universal, thus, under selection we could only test for moderation by the presence or absence of selection for individual risk and selection for environmental or ecological risk. We could not consider personal experience as a motivation as there were no studies in which this was measured or was able to be coded. Thus moderation tests of mentor motivations were conducted separately for presence or absence of civic duty and for professional development as motivation.

Only the tests of inclusion of other interventions with mentoring included all 46 studies. Other moderator analyses were limited by whether coders could determine

whether the moderating factor was present or absent. The analysis of whether motivation by civic duty significantly moderated effect sizes included 36 studies, which was the smallest number of studies in any of the moderation analyses.

To conduct the moderated analyses we utilized all studies across the four outcomes to calculate an overall effect size by moderator condition (i.e., the mean of all effect sizes reported in each study). This was done because of the limited number of studies for testing moderators available even if examined for each outcome separately. We also reasoned that the interest was in testing moderation of mentoring for studies of delinquency and/or the related outcomes rather than for each specific outcome. That is, this meta-analysis is focused on youth at-risk for delinquency from the view that the four outcomes are related in sharing risk factors and likely impact of mentoring features. This approach has been used in other meta-analyses where multiple outcomes are of interest (see DuBois et al., 2011). In addition, given the power strain moderation analyses can impose on data sets limited in size like this one, as has been done by others we utilized a *p* level of .05 (one-tailed test). This standard was also used for these analyses because in each case we expected larger effects if the moderator was present, and the specific order of the levels of the moderator was not at issue. This is equivalent to a two-tailed $p < .10$ which has been justified given the power challenges for moderation effects (Wilson & Lipsey, 2007).

We tested for moderation with meta-regression analysis using the *rma* function in the *metafor* package in *R* (Viechtbauer, 2010). Each meta-regression analysis employed a random effects model that included terms for the moderator under consideration and a term representing whether the study was a randomized design or a quasi-experimental trial. The significance tests are one-tailed Z-tests.

We also conducted sensitivity analyses to assess the effects on conclusions of changes made to the inputs of an analysis (Morgan & Henrion, 1990). Accordingly, we conducted analyses to determine (1) the consistency of effect sizes obtained with different outcome variables, and (2) the consistency of outcomes within different levels of moderated analyses.

4 Results

4.1 MAIN EFFECT META-ANALYSES RESULTS

In the updated review's first phase of the literature search we identified additional studies to accumulate a total of 164 studies that were further evaluated for basic criteria for outcome and intervention type. Of these studies, 58 (34%) were determined to have none of the target outcomes. The remaining 107 were subjected to further scrutiny in order to determine their methodological suitability for the meta-analysis. Of these 53 (33%) had research designs that did not meet minimum quality standards for inclusion and 6 (4%) did not provide sufficient information for calculating effect sizes related to the outcomes in question This left 46 (28%) studies that were included in the quantitative review. The 118 excluded studies can be found in Table 7.

Table 3 provides details on the 46 studies selected for the meta-analysis, including citation, sample characteristics, design type, component and intervention information obtained for moderation analyses, and basic findings. Of the 46 studies included, 27 were randomized controlled trials and 19 were quasi-experimental studies involving non-random assignment, but with matched comparison groups as was described above. Twenty-five studies reported delinquency outcomes, 25 reported academic achievement outcomes, 6 reported drug use outcomes, and 7 reported aggression as an outcome.

Prior to calculating the mean effect size, we evaluated the heterogeneity of study effect sizes using multiple homogeneity measures, standard errors, and associated probability levels, including Cochrane's *Q,* and I^2 (Higgins, Thompson, Deeks & Altman, 2003). Cochrane's *Q* is an indicator of heterogeneity that is distributed as a *chi-square*. Significant values of *Q* indicate heterogeneity. The degree of heterogeneity can be seen in the I^2 statistics. This indicates the approximate proportions of variance across compared studies that are due to heterogeneity of effects.[1]

1 In a sensitivity analysis we tested for influence of studies with multiple outcomes on effects and found that the effect sizes in studies with single outcomes (SMD = 0.27, 95% CI = 0.12 - 0.41) were slightly but not significantly higher than the effect sizes in studies with multiple outcomes (SMD = 0.22, 95% CI = 0.07 - 0.38). Cross-tabulation of multiple outcomes by moderator variables revealed a single significant difference. Studies with a single outcome were more likely to have selected for

We inspected forest plots of the effects and confidence intervals to explore for potential outlying studies. Our procedure was, after identifying possible outlying studies we repeated the meta-analyses, in order to determine whether removal of up to five outlying studies would reduce or eliminate the heterogeneity.

As can be seen in Table 4, heterogeneity of effects was substantial for delinquency and academic achievement. Also, examination of forest plots and re-analysis with removal of outlying studies did not reduce appreciably the heterogeneity of effects of mentoring for either delinquency or academic achievement. It seems evident there is substantial heterogeneity among studies in effects for delinquency and academic achievement.

In order to assist in understanding the heterogeneity in effect sizes, we conducted an analysis to determine whether the effect sizes differed substantially between randomized controlled trials (RCTs) and quasi-experimental designs. Using meta-regression with study design as the predictor, we found that although effect sizes were numerically larger in RCTs for all outcomes except drug use, none of these differences was statistically significant (Hedges & Pigott, 2004, formulas 11-12, p. 432).

For each outcome we calculated an average effect size and 95% confidence interval and a related *Z* statistic. To facilitate interpretation, we scaled all outcomes so that positive effect sizes represent effects in the desired direction, i.e., lower delinquency, aggression and drug use, higher academic achievement or lower school failure. Table 4 reports the results for the meta-analysis for each of the four studied outcomes.

Delinquency

As can be seen in Table 4 the 25 studies with a delinquency outcome yielded an average effect size of SMD = .21. (Range: -0.25 to 1.73; 95% confidence interval 0.17 to 0.25; $p < .01$). Heterogeneity was substantial as indicated by I2 of 99.3% ($Q(24) = 3297.64$, $p < .01$). Examination of a funnel plot for delinquency revealed some asymmetry involving the three studies with the largest effect sizes, and an Egger test confirmed the presence of asymmetry (bias = 6.79, $t(23) = 2.74$, $p < .05$). We conducted a sensitivity analysis by removing these studies and repeating the meta-analysis. The difference was very slight. With the full sample, the SMD from the random effects model was 0.21 ($p < .001$; $\tau 2 = .008$). With the reduced sample the SMD from the random effects model was 0.19 ($p < .001$; $\tau 2 = .008$). Finally, we applied the trim and fill method (Duval & Tweedie, 2000) to account for publication bias in the random effects estimate. The result was an estimated effect of 0.18 ($p < .001$; $\tau 2 = .009$).

environmental or ecological risk than were studies that reported multiple outcomes, $\chi 2 (1, N=36) = 3.94$, $p < .05$

Aggression

As can be seen in Table 4 a random effects model of the seven studies with Aggression outcome yielded an average weighted effect size of *SMD* = .29 (Range: -0.05 to 0.95; 95% confidence interval: -0.03 to 0.62, *ns*). The funnel plot for Aggression revealed no asymmetry and the Egger test confirmed this impression (bias = -1.41, t (5) < 1, *ns*).

Drug Use

As can be seen in Table 4 a random effects model of the six studies with Drug Use outcome yielded an average weighted effect size of *SMD* = .16 (Range: -0.13 to 0.18; 95% confidence interval*:* 0.04 to 0.29, p = .05). On drug use, there appeared to be funnel plot asymmetry due to the single negative effect, but the Egger test did not find evidence of bias (bias = 16.41, t (4) < 1, *ns*). Removal of this effect in a sensitivity analysis resulted in stronger combined effect (Full sample: *SMD* = .16, p = .05, τ^2 = .04; Reduced sample: *SMD* =. 19, p < .001, τ^2 = .0002).

Academic Achievement

As can be seen in Table 4 the 25 studies with Academic Achievement outcome yielded an average effect size of *SMD* = .11 (Range: -0.04 to 1.45; 95% confidence interval: 0.03 to 0.31). On academic achievement, graphical examination suggested that there might be funnel plot asymmetry due to three studies with large effect sizes. Removal of these effects in a sensitivity analysis resulted in a weaker, but still significant combined effect (Full sample: SMD=.11, p < .0001, τ^2 = .006; Reduced sample: *SMD* =.05, p < .01, τ^2 = .005). An Egger test of bias found no evidence of bias with the full sample (bias=4.55, t (23) = 1.65, p = .11).

Average Effect

Table 4 also reports the average effect, which was used for the moderation analysis. The 46 studies yielded an average effect size of *SMD* = .18 (Range: -0.21 to 1.70; 95% confidence interval: 0.15 to 0.21).

We also created forest plots for each outcome to show the variation in individual studies about the aggregate effect size. These are the effect sizes from inverse variance weighted random effects models. These are provided, with accompanying statistics, in Figures 1-4, corresponding to Delinquency, Aggression, Drug Use, and Academic Achievement respectively. Across the four outcomes the pattern is one of relatively consistent direction and size of effect sizes within a given outcome, but with a few studies showing confidence intervals that include zero or negative effects for each outcome.

The patterns of effect sizes and the Forest Plots suggest the average effect sizes represent robust estimates of mentoring on each outcome. The aggregate effect size estimates, although modest, are all positive.

4.2 MODERATOR ANALYSES

We conducted analyses to determine whether the effects of the mentoring interventions varied by four key aspects of intervention design and implementation and of four key processes theorized as avenues for mentoring effects:

1) selectivity in inclusion (high individual risk, high environmental risk, or no such selectivity)
2) whether or not mentoring is a stand-alone approach in that study or was undertaken along with a) some other major intervention components or b) some relatively minor add-ons
3) the motivation of the mentors in participating (civic duty, professional development, own experience)
4) the extent to which quality of work and fidelity were assessed or emphasized.
5) explicit attention to presence of four key processes: modeling/identification promotion, emotional support, advocacy, and teaching.

As noted earlier we combined across outcomes and for these analyses given the constrained sample sizes and used a test as all moderator analyses tested the null hypothesis that the effect of the moderator was zero, regardless of which level of the moderator was coded "1" and which was coded "0". To check on the validity of combining across outcomes we tested for bias in effects due to this aggregation (e.g. effects are limited to one outcome or heavily dependent on specific outcome). To do so we conducted two sets of sensitivity analyses. For the first set of analyses, we employed Hedges and Pigott's (2004, formulas 11-12, p. 432) method for contrasting group mean effect sizes in meta-analysis to contrast effect sizes from studies reporting delinquency outcomes against those reporting each outcome against those reporting on the other three outcomes. These results produced no evidence that effect sizes differed substantially by any given outcome, which would mean moderation relations were not due to a true relation with only a single outcome, Z (delinquency-aggression) = -0.17, ns; Z (delinquency-drug use) = 1.61, ns; Z (delinquency-academic) = 1.77, ns; Z (aggression-drug use) = 0.74, ns; Z (aggression-academic) = 0.81, ns; and Z (academic-drug use) = -0.07, ns. We also coded outcomes of each study according to the outcome variables used (e.g., 1-4 = Delinquency, Aggression, Drug Use, Academic Achievement). We then cross-tabulated these codes with categorical scores for whether a given moderator could be coded. No significant results were obtained. Only one moderator, professional development as a motivation for mentoring, showed any such tendency, with a marginally higher than expected frequency by outcome (for academic achievement) $\chi 2$ (5, n=36) = 11.05, $p < .05$. These results suggested to us sufficient confidence that moderation analyses collapsed across outcomes would be not biased or

misrepresenting an overall relation for mentoring programs. In combination with the practical consideration of sample size limitations we judged this an appropriate way to serve the goals of the review with the available studies.

We tested for moderation using two methods. First, we calculated meta-analysis statistics separately by levels of the moderators (Hunter & Schmidt, 2004, p. 402). Table 6 reports the standardized mean difference effect sizes by levels of each moderator, the number of studies in each level of the moderator, and the lower and upper limits of the 95% confidence intervals for each random effect estimate. Table 6 also reports the moderator effect estimates, standard errors, and significance tests from the meta-regression analyses described above.

As can be seen in Table 6 there was significant moderation for Motivation for Mentoring but not for other program organization and implementation features. We provide plots for Mentor Motivation in Figure 5. As can be seen in Figure 5 effects were larger when mentor motivation was based in professional development.

In regard to key processes of mentoring interventions, there was evidence of significant moderation by the presence of two component processes in mentoring: Advocacy and Emotional Support (See Table 6). The results are illustrated in Figure 6. Stronger effects were observed when Emotional Support and Advocacy were components of mentoring than when these components were not present. Figure 6 suggests that stronger effects were observed when teaching was a component of mentoring, but the meta-regression that included a term for research design did not return significant evidence of moderation.

5 Conclusions

This review of the methodologically adequate studies released between 1970 and 2011 and focused on primarily United States population testing mentoring for high-risk youth found positive effects for delinquency and for three other associated outcomes: aggression, drug use, and academic performance. These findings suggest mentoring is beneficial for at-risk youth to reduce delinquency, aggression, substance use, and to improve academic functioning. In addition, we found that the size of the effects varied by some key features, which include Mentor's motivation for being a mentor (those with interest in professional development had large effects) and for two of four theorized key processes were part of the mentoring effort (Advocacy and Emotional Support, with strong suggestion for Teaching). While showing these overall effects, for each outcome and among the studies with the beneficial features, there was substantial variation in effect sizes.

The effects are significantly different from zero for all four outcomes. However, all were modest in size (ranging from .11 for Academic Achievement to .16 for drug use, .21 for delinquency and .29 for aggression). These effect sizes are comparable to other interventions aimed at high-risk youth for each outcome.

These results suggest mentoring, at least as represented by the included studies, has positive effects for these important public health problems with those at risk for delinquency. As this portion of the population can be of particular interest given the problems their elevated risk for not just delinquency but many other areas of functioning, the evidence of mentoring having significant effects, even if modest in size, suggest it could be part of the strategies to try to prevent actual engagement in delinquency and drug use and to curtail or prevent aggression and poor academic achievement (Tolan & Gorman-Smith, 2003). In addition, there was substantial heterogeneity in effect size across programs for each outcome suggesting there may be more substantial benefits that could be gained when mentoring is organized in ways that maximize those features associated with larger effects.

However, there were several limitations of the available literature that preclude statements about what makes mentoring most effective or what accounts for benefits. Perhaps most notably, the collected set of articles is remarkably limited in

describing the actual program activities, what was expected and not among a range of potential mentoring activities, and how key implementation features were organized, trained, and/or assessed for competence and fidelity. Unfortunately this state of reporting detail and completeness does not seem to be improving such that more recent publications are clearly more informative.

This longstanding concern is part of what prompted the formulation of key processes and attempted coding of these in this review. As we noted in the introduction and as we attempted to code, there are key characteristics thought to distinguish mentoring from other helping relationships and to be the basis for benefits. Therefore, these qualities should be common across studies and their quality relate to effect size. However, for a significant portion of studies description of the intervention content, organization, and/or implementation was insufficient to code one or more of these important characteristics. This state of the reporting of details about intervention constrained sensitivity of our moderation analyses and completeness of the comparisons for the body of research considered here.

The notable lack of adequate reporting of specific components, implementation procedures and adherence, and measurement of targeted processes to permit comparison on these important features is seen as a major impediment to advancing knowledge about the value of this popular approach to youth intervention. It may be that full potential of the approach is not being achieved, as what may improve effects is difficult to discern. Importantly, there is limited ability to meta-analytically determine what characteristics of mentoring programs and which approaches are most advantageous and might provide direction for more effective programs. Thus, we have limited ability to suggest specific priorities for further study.

We were able to conduct some moderator analyses despite these limitations. The results for tests of several features of organization and implementation of mentoring suggest that effects were larger when mentors were motivated to participate by interest in advancing their professional careers. This is an important finding as most mentoring is undertaken as voluntary activity. In some cases the mentoring may help a mentor by fulfilling requirements at work, as an entry level position toward a professional staff position, or by enabling experience that can make them a more attractive candidate for educational or occupation opportunities. While beyond the scope of this review, the results may also raise questions about the presumption that mentoring should not be done other than as a voluntary activity.

Although the review focused on selective and indicated populations (those with risk characteristics or already exhibiting delinquency as a basis for inclusion) we did not find moderation by whether inclusion depended on individual risk characteristics or environmental or other-than-individual characteristics. While we are duly cautious about interpreting these null effects, the finding may suggest that either approach may be viable for effective targeting.

We also did not find effect differences by whether or not other interventions were included with mentoring or mentoring was part of a multi-component intervention than when it was offered on its own. This leaves open whether or not the effect when other interventions are present is attributable to mentoring but does suggest that mentoring, at least as represented in these collected studies, has effects apart from those attributable to other interventions. Within the overall concern about the quality of information about mentoring programs there is much need to consider designs that might consider mentoring singularly and as part of a package or in comparison to other singular interventions. This could not only help clarify the relative importance of other components but also the relative value in comparison to other interventions that might be alternatives. As issues such as cost effectiveness, ease of training and implementation, and sustainability come into consideration, such information is increasingly important.

Similarly, we did not find differences by whether or not extent and fidelity of implementation of expected activities and program features was measured (?). While what comprises a mentoring program to test fidelity against is in some cases not clear, the impression from the limited number of studies we could code for this is that this field is behind others in such design and evaluation considerations. As with the other factors noted here, more attention to this would likely improve understanding and efficiency of program advancement.

Moderation tests of four key processes found to be mentioned frequently in the literature and in description of some programs found that at least two matter in regard to effects. Programs that included emphasis on emotional support and those that emphasized advocacy for the recipient had larger effects. While teaching and modeling/identification did not significantly relate to effect size, there was some suggestion these may be worthwhile foci of attention in mentoring design. Perhaps with more studies that could be coded and more attention to documentation of such processes, the role of these four processes can be better delineated. The present results suggest programs might want to ensure emotional support from the mentor is emphasized but also methods and opportunities to advocate could also be helpful. Our results in regard to the latter are consistent with those reported by DuBois et al. (2011) for mentoring in general when measured across many outcomes.

These findings are consistent with prior meta-analyses that overlap in focusing on mentoring. As reported by Lipsey and Wilson (1998) and DuBois et al. (2002, 2011) these analyses suggest general support for mentoring for intervention related to delinquency and closely associated outcomes. However, as those analyses found, the information obtainable about the “inside” of these interventions termed mentoring is limited. Thus, the conclusions to be drawn must remain very sketchy about what it is that makes mentoring effective. This persistent characteristic of the field undercuts ability to recommend it for use, as it is not clear what should be

recommended. Further, while the positive effects suggest promise, the lack of standard types of information and formal approaches to documentation that characterizes the best studies in most areas of behavioral intervention seriously impedes incremental progress in best practices. Thus, while consistent with prior findings, there seems to be little additional certainty of the nature of mentoring and information to guide further development, sound training and management of the programming, and adequate tracking of effects to activities, staffing, and other features. Unfortunately this seems to be qualitatively the same state of need as was identified in our consideration of mentoring in a review of violence prevention 14 years ago (Tolan & Guerra, 1994). This is not the case for most areas of delinquency intervention.

This lack of progress and lack of attention to intervention design features and program characteristics is particularly of note because mentoring is one of the most common and most favored approaches for prevention of risk and youth development. It is also one with considerable presence in the scientific literature. While of the 164 studies located only 46 met criteria for inclusion, this does not mean the other 118 were of no value for informing science. Yet, after reviewing these we do note they are not marked by more detailed attention to these conventions of design and reporting that have helped advance prevention and intervention capabilities for other approaches. Given the prominence of mentoring in attempts to address these critical public health and youth problems, such a lack of systematic attempts to unpack mentoring and to understand it within a conventional framework for evaluating interventions is surprising. It is also striking that funding and promotion of these efforts proceed without more stringent evaluation, including more careful identification of population of interest, inclusion criteria, skills and training of providers, content and theorized processes of component effects, fidelity tests, and implementation levels for intent to treat.

Thus, we can only suggest some tentative and general statements about what might affect mentoring impact. Perhaps the more striking statement to be made is that despite its popularity and the apparent benefits it provides, there is little understanding of just what makes an intervention mentoring and what about such labeled interventions is related to benefits derived. Perhaps most fundamentally the co-occurring popularity and the general promise of these findings point to the critical need for concerted efforts for substantial and probably large-scale evaluations. These are needed to efficiently provide more clear and directing information about what about mentoring is the reason positive effects are derived. In particular it may be that the promise suggested in the modest effect sizes yielded here is only a base estimation of potential benefit. Similarly, the suggestion that some design features and some emphases are related to larger effects may only point to the potential gain that could come from more careful and concerted formalization of intervention evaluation.

6 Plans for updating the Review

The review will be updated every 5 years.

7 Review References

7.1 REFERENCES FOR REPORT

Achenbach, T. M. (1991). *Manual for Child Behavior Checklist/ 4-18 and 1991 Profile*. Burlington, VT: University of Vermont, Dept. of Psychiatry

Aos, S., Lieb, R., Mayfield, J., Miller, M. & Pennucci, A. (2004). *Benefits and costs of prevention and early intervention programs for youth*. Olympia: Washington State Institute for Public Policy.

Bailey, K. R. (1987). Inter-study differences: How should they influence the interpretation and analysis of results? *Statistics in Medicine, 6*, 351–358.

Beier, S. R., Rosenfeld, W. D., Spitalny, K. C., Zansky, S. M., & Bontempo, A. N. (2000). The potential role of an adult mentor in influencing high-risk behaviors in adolescents. *Archives of Pediatrics & Adolescent Medicine, 154*, 327-331.

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine, 19*, 3127-3131.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

Deutsch, N., & Spencer, R. (2009). Capturing the magic: Assessing the quality of youth mentoring relationships. *New Directions in Youth Development: Theory, Practice and Research*, 121, 47-70.

DuBois, D. L., Holloway, B. E., Valentine, J. C., & Cooper, H. M. (2002). Effectiveness of mentoring programs for youth: A meta-analytic review. *American Journal of Community Psychology, 30*, 157-197.

DuBois, d. L., & Karcher, J. J. (2005). Youth mentoring: Theory, research, and practice. In D. L. DuBois & M. J. Karcher (Eds), Handbook of youth mentoring pp2-12). Thousand Oaks, CA: Sage

Duval S & Tweedie R (2000), A nonparametric "Trim and Fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*, 89–98.

Egger, M., Smith, G. D. S., M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629-634.

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.

Hall, J. C. (2003). *Mentoring and young people: A literature review*. Glasgow, Scotland: The SCRE Centre, University of Glasgow.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107-128.

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

Hedges, L. V., and Pigott. T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods 9*, 426–445.

Hieronymus, A. N., Lindquist, E. F. & Hoover, H. D. (1986).*Iowa Tests of Basic Skills: Early Primary Battery*. Chicago: Riverside

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327,* 557-560.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings (2nd ed.)*. Newbury Park: Sage Publications.

Jekielek, S. M., Moore, K. A., Hair, E. C., & Scarupa, H. J. (2002). *Mentoring: A promising strategy for youth development* (research brief). Washington, DC: Child Trends.

Lipsey, M. W., & Wilson, D. B. (1998). Effective intervention for serious juvenile offenders: A synthesis of research. In R. Loeber & D. Farrington (Eds.), *Serious and violent juvenile offenders: Risk factors and successful interventions*. (pp. 313-341). Thousand Oaks, CA: Sage.

Landis J.R., & Koch G.G. (1977). The measurement of observer agreement for categorical data, *Biometrics, 33*, 159-174.

McCord, J. (1978). A thirty-year follow-up of treatment effects. *American Psychologist, 33*(3), 284-289.

McCord, J. (1979). Following up on Cambridge-Somerville: Response to Paul Wortman. *American Psychologist, 34*(8), 727.

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.

MENTOR/National Mentoring Partnership. (2006). *Mentoring in America 2005: A snapshot of the current state of mentoring*. Alexandria, VA: Author.

MENTOR/National Mentoring Partnership. (2009). *Elements of effective practice for mentoring* (3rd ed.). Alexandria, VA: Author. Retrieved from http://www.eric.ed.gov/PDFS/ED512172.pdf

Morgan, G., & Henrion, M. (1990). *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge: Cambridge University Press.

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

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

Raudenbush, S. W. (1994). Random effects models. In: H. Cooper & L. V. Hedges (eds.), *The handbook of research synthesis*. New York: Russell Sage Foundation.

Reynolds, C. R., & Kamphaus, R. W. (1992). *Behavior Assessment System for Children\ (BASC)*. Circle Pines, MN: American Guidance Service.

Rhodes, J. E. (2002). *Stand by me: The risks and rewards of mentoring today's youth*. Cambridge, MA: Harvard University Press.

Rhodes, J. E., Bogat, G. A., Roffman, J., Edelman, P. & Galasso, L. (2002). Youth mentoring in perspective: Introduction to special issue. *American Journal of Community Psychology, 30*, 149-155.

Rhodes, J. E., Spencer, R., Keller, T. E. Liang, B. & Noam, G. (2006) A model for the influence of mentoring relationships on youth development. Journal of Community Psychology, 34(6), 691-707

Roberts, H., Liabo, K., Lucas, P., DuBois, D., & Sheldon, T. A. (2004). Mentoring to reduce antisocial behavior in childhood. *British Medical Journal, 328*, 512-514.

Tolan, P.H. (2002). Crime prevention: Focus on youth. In J.Q. Wilson & J. Petersilia (Eds.), *Crime* (pp. 109-128). Oakland, CA: Institute for Contemporary Studies Press.

Tolan, P. H., & Brown, C. H. (1998). Methods for evaluating intervention and prevention efforts. In. P. K. Trickett & C. Schellenbach (Eds.), *Violence against children in the family and the community* (pp. 439-464). Washington, DC: American Psychological Association.

Tolan, P.H., & Gorman-Smith, D. (2003). What violence prevention can tell us about developmental psychopathology. *Developmental and Psychopathology*.

Tolan, P. H., & Guerra, N. G. (1994). *What works in reducing adolescent violence: An empirical review of the field*. Boulder: University of Colorado, Center for the Study and Prevention of Youth Violence.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. Journal of Statistical Software, 36(3), 1-48.

Wilson, S.J., & Lipsey, M.J. (2007). Effectiveness of school-based intervention programs on aggressive behavior: Update of a meta-analysis. *American Journal of Preventive Medicine, 33*(Suppl. 2), S130–S143.

7.2 REFERENCES FOR STUDIES INCLUDED IN SYSTEMATIC REVIEW

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.

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.

Anderson, P. S. (1977). *Volunteers in criminal justice: A three-year evaluation of the Clark County, Washington program*. Vancouver, WA: Health and Welfare Planning Council.

Aseltine, R. H., Dupre, M. & Lamlein, P. (2000). Mentoring as a drug prevention strategy: An evaluation of Across Ages. *Adolescent and Family Health, 1(1),* 11-20.

Barnoski, R. (2002). Preliminary findings for the Juvenile Rehabilitation Administration's Mentoring Program. Olympia, WA: Washington State Institute for Public Policy. (WSIPP Doc # 02-07-1202)

Berger, R. J. & Gold, M. (1978). An evaluation of a juvenile court volunteer program. *Journal of Community Psychology, 6*, 328-333.

Bernstein, L., Rappaport, C., Olsho, L., Hunt, D.& Levin, M. (2009) Impact evaluation of the US Department of Education's student mentoring program. National Center for educational evaulation and regional assistance.

Blechman, E. A., Maurice, A., Buecker, B., Helberg, C. (2000). Can mentoring or skill training reduce recidivism? Observational study with propensity analysis. *Prevention Science, 1(3)*, 139-155.

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

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

Cavell, T. A., & Hughes, J. N. (2000). Secondary prevention as context for assessing change processes in aggressive children. *School Psychology, 38(3)*, 199-235.

Clarke, L. (2009) Effects of a school-based adult mentoring intervention on low income, urban high school freshman judged to be at risk for drop out: a replication and extension. PhD Dissertation, The State University of New Jersey, United States--New Jersey.

Converse, N. & Lignugaris/Kraft, B. (2009) Evaluation of a school-based mentoring program for at-risk middle school youth. Remedial and special education, 30

Davidson, W. S., II (1976). The diversion of juvenile delinquents: An examination of the processes and relative efficacy of childhood advocacy and behavioral contracting. (Doctoral dissertation, University of Illinois at Urbana-Champaign, 1976). *Dissertation Abstracts International, 37(01),* 601. (University Microfilms No. AAT 7616113)

Davidson, W. S., II. & Redner, R. (1988). 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: A Casebook for Practitioners* (pp. 123-137). Washington, D.C.: American Psychological Association.

Davidson, W. S., II., Amdur, R. L., Mitchell, C. M., & Redner, R. (1990). *Alternative Treatments for Troubled Youth: The Case of Diversion from the Justice System*. New York: Plenum Press.

Davidson, W. S., Seidman, E., Rappaport, J., Berck, P. L., Rapp, N. A., Rhodes, W., & Herring, J. (1977). Diversion program for juvenile offenders. *Social Work Research and Abstracts, 13(2)*, 40-49.

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). Washington D.C: American Psychological Association.

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

Dicken, C., Bryson, R., & Kass, N. (1977). Companionship therapy: A replication of experimental community psychology. *Journal of Consulting and Clinical Psychology, 45(4),* 637-646.

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

Fo, W. S. O. & O'Donnell, C. R. (1972). *The Buddy System model: Community-based delinquency prevention utilizing indigenous nonprofessionals as behavior change agents.* Washington, DC: U.S. Department of Health, Education & Welfare Office of Education. (ERIC Document Reproduction Service No. 074-394)

Fo, W. S. O. & O'Donnell, C. R. (1979). The Buddy System: Relationship and contingency conditions in a community intervention program for youth with nonprofessionals as behavior change agents. In J. S. Stumphauzer (Ed.), *Progress in behavior therapy with delinquents* (pp.302-316). Springfield, IL: Charles C. Thomas.

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

Grant, C (2010). Grounded in your culture: the hidden key to promoting academic achievement among african american adolescent males. PhD Dissertation Capella University, United States – Minneapolis.

Grossman, J. B. & Tierney, J. P. (1998). Does mentoring work? An impact study of the Big Brothers Big Sisters Program. *Evaluation Review 22(3),* 403-426.

Grossman, J. B., & Rhodes. J. E. (2002). The test of time: Predictors and effects of duration in youth mentoring relationships. *American Journal of Community Psychology, 30(2),* 199-219.

Hanlon, T. E., Bateman. R. W., Simon, B. D., O'Grady, K. E., & Carswell, S. B. (2002). An early community-based intervention for the prevention of substance abuse and other delinquent behavior. *Journal of Youth and Adolescence 31(6),* 459-471.

Harmon, M. A. (1995). Reducing drug use among pregnant and parenting teens: A program evaluation and theoretical examination. (Doctoral dissertation, University of Maryland at 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 dissertation, Portland State University, 1998). *Dissertation Abstracts International, 59(05),* 1517A.

Herrera, C., Grossman, J., Kaugh, T., Feldman, A. & McMaken, J. (2007) Making a difference in schools: The big brothers big sisters school based mentoring impact study. Philadelphia: Public/Private Ventures.

Holt, L., Bry, B. & Johnson, V.(2008) Enhancing school engagement in at-risk, urban minority adolescents through a school-based, adult mentoring intervention. Child & Family Behavior Therapy, 30(4), 297-318

Johnson, A. (1999). *Sponsor-a-Scholar: Long-term impacts of a youth mentoring program on student performance*. Retrieved November 12, 2003 from http://www.cmwf.org/programs/youth/johnson_sp-a-sch_355.

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.

Karcher, M. (2008) The study of mentoring in the learning environment (SMILE): A randomized evaluation of the effectiveness of school-based mentoring. Prevention Science, 9, 99-113

Keating, L. M. (1996). Mentorship as prevention of juvenile delinquency (Doctoral dissertation, California School of Professional Psychology, San Diego, 1996). *Dissertation Abstracts International, 57(04),* 2869. (University Microfilms No. AAT 9626966)

Keating, L. M., Tomishima, M. A., Foster, S., & Alessandri, M. (2002). The effects of a mentoring program on at-risk youth. *Adolescence, 37(148),* 717-734.

Kelley, T. M. (1973). I. Student volunteer effectiveness in a delinquency prevention experiment. II. Validation of a selection device for volunteer probation officers. (Doctoral dissertation, Wayne State University, 1972). *Dissertation Abstracts International, 33(11),* 140.

Kelley, T. M., Kiyak, H. A., & Blak, R. A. (1979). The effectiveness of college student companion therapists with predelinquent youths. *Journal of Police Science and Administration, 7(7),* 186-195.

Kemple, J. J., & Scott-Clayton, J. (2004). *Career academies: Impacts on labor market outcomes and educational attainment*. New York: Manpower Demonstration Research Corporation.

Ku, R. & Blew, C. H. (1977). *A University's approach to delinquency prevention: The Adolescent Diversion Project: An exemplary project*. Rockville, MD: National Institute of Justice. (NCJRS Document Reproduction Service No. 38510)

Lattimore, C. B., Mihalic, S. F., Grotpeter, J. K., & Taggart, R. (1998). *Blueprints for violence prevention, book four: The Quantum Opportunities Program*. Boulder, CO: Center for the Study and Prevention of Violence.

LoSciuto, L., Rajala, A. K., Townsend, T. N., & Taylor, A. S. (1996). An outcome evaluation of Across Ages: An intergenerational mentoring approach to drug prevention. *Journal of Adolescent Research, 11(1),* 116-129.

Maxfield, M., Schirm, A., & Rodriguez-Planas, N. (2003, August). *The Quantum Opportunity Program demonstration: Implementation and short-term impacts*. MPR Reference No. 8279-093. Washington, D.C.: Mathematica Policy Research, Inc.

McCord, J. (1978). A thirty-year follow-up of treatment effects. *American Psychologist, 33(3),* 284-289.

McCord, J. (1979). Following up on Cambridge-Somerville: Response to Paul Wortman. *American Psychologist, 34*(8), 727.

Moore, R. H. & Levine, D. (1974*). Evaluation research of a community-based probation program*. Lincoln, NE: University of Nebraska at Lincoln, Department of Psychology. (ERIC Document Reproduction Service No. 132497)

Moore, R. H. (1987). Effectiveness of citizen volunteers functioning as counselors for high-risk young male offenders. *Psychological Reports, 61*, 823-830.

Newton, F. R. (1994). A study of the effectiveness of using collegiate mentors to reduce violent behavior, improve self concept, and increase academic success in an urban middle school (at risk). (Doctoral dissertation, The College of William and Mary, 1994). *Dissertation Abstracts International, 55(06-A),* 1440.

O'Donnell, C. R., Lydgate, T., & Fo, W. S. O. (1979). The Buddy System: Review and Follow-Up. *Child Behavior Therapy, 1(2),* 161-169.

Polit, D., Kahn, J., & Stevens, D. (1985). *Final impacts from Project Redirection: A program for pregnant and parenting teens*. New York: Manpower Demonstration Research Corporation.

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

Reyes, O., & Jason, L. A. (1991). An evaluation of a high school dropout prevention program. *Journal of Community Psychology, 19,* 221-230.

Rhodes, J., J. Grossman, & Resch, N. (2000). Agents of change: Pathways through which mentoring relationships influence adolescents' academic adjustment. *Child Development, 71(6),* 1662-1671.

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(129),* 145-158.

Schinke, S. P., Cole, K. C., & Poulin, S. R. (2000). Enhancing the educational achievement of at-risk youth. *Prevention Science, 1(1),* 51-60.

Seidman, E., Rappaport, J., & Davidson, W. S., II (1980). Adolescents in legal jeopardy: Initial success and replication of an alternative to the criminal justice system. In R. R. Ross, & P. Gendreau (Eds.), *Effective correctional treatment* (pp. 103-123). Toronto: Butterworth.

Taylor, A. S., LoSciuto, L., Fox, M., Hilbert, S. M., & Sonkowsky, M. (1999). The mentoring factor: Evaluation of the Across Ages' intergenerational approach to drug abuse prevention. In V.S. Kuehne (Ed.), *Intergenerational programs: Understanding what we have created* (pp. 77-99). New York: Haworth Press.

Taylor, A. S., LoSciuto, L., Fox, M., Hilbert, S. M., & Sonkowsky, M. (1990). The mentoring factor: Evaluation of the across ages' intergenerational approach to drug abuse prevention. *Child Youth Services, 20(1/2),* 77-99.

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 (Master’s thesis, Stephen F. Austin State University, 1996). *Masters Abstracts International*, *35*(05), 1133.

7.3 REFERENCES FOR CODED STUDIES EXCLUDED FROM SYSTEMATIC REVIEW ANALYSES

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

Ahrens, K., Richardson, L., Lozano, P., Fan, M. & Dubois, D (2008). Foster care youth with adult mentors during adolescence have improved adults outcomes.Pediatrics, 121 (2), 246-252

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)

Barron-McKeagney, T., Woody, J. D., & D'Souza, H. J. (2001). Mentoring at-risk Latino children and their parents: Impact on social skills and problem behaviors. *Child Adolescent Social Work, 18(2)*, 119-136.

Bellamy, N. D., Springer, U. F., Sale, E. W., & Espiritu, R. C. (2004). Structuring a multi-site evaluation for youth mentoring programs to prevent teen alcohol and drug use. *Journal of Drug Education, 34(2)*, 197-212.

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

Blinn-Pike, L., Kuschel, D., McDaniel, A., Mingus, S., & Mutti, M. P. (1998). The process of mentoring pregnant adolescents: An exploratory study. *Family Relations, 47(2)*, 119-127.

Bracy, D.. Female adolescent's response to self-efficacy through an intervention mentoring program. Ph.D. dissertation, TUI University, United States -- California. Retrieved July 2, 2010, from Dissertations & Theses: Full Text.(Publication No. AAT 3317639).

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

Campos-Torres, C., Phinney, J., Perez-Brena, N., Chami, K., Ornelas, B., Nemanim, L., Kallemeyn, D., Mihecoby, A.& Ramirez, C (2009). A mentor-based teargeted intervention for high-risk latino college freshman: A pilot study. Journal of Hispanic Higher Education, 8 (2), 158-178

Cave, G., & Quint, J. Manpower Demonstration Research Corporation. (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)

Cheng, T., Haynie, D., Brenner, R., Wright, J., Chung, S.& Simon-Morton, B (2008). Effectiveness of a mentor-implemented, violence prevention intervention for assault-injured youths presentingn to the emergency department: results of a randomized trial. Pediatrics, 122, 938-946.

Colley, H. (2003). Engagement mentoring for "disaffected" youth: A new model of mentoring for social inclusion. *British Educational Research Journal, 29(4)*, 521-542.

Colley, H. (2003). Engagement mentoring for socially excluded youth: Problematising an "holistic" approach to creating employability through the transformation of habits. *British Journal of Guidance & Counselling, 31(1)*, 77-99.

Colson, S., Godsey, R., Mayfield B., Nash, W. R., & Borman, C. (1978). Evaluating career education for gifted and talented students. *Journal of Research and Development in Education, 12*, 51-62.

Conduct problems prevention research group (2007). Fast track randomized controlled trial to prevent externalizing psychiatric disorders: findings from grades 3 to 9. Jounral of American Acadamy of Child and Adolescencent Psychiatry, 46(10), 1250-1262

Cummings, L. (2010) evaluating the influence of participation in a diverse high school based group mentoring program. PhD Dissertation--Georgia State University - Atlanta, GA

Dance, L .J. (2001). Shadows, mentors, and surrogate fathers: Effective schooling as critical pedagogy for inner-city boys. *Sociological Focus, 34(4)*, 399-415.

Dappen, L., & Isernhagen, J. C. (2002). TeamMates: A model to support mentoring in rural schools. *Journal of Research in Rural Education, 17(3)*, 154-161.

Davis, T., & Haney, J. R. (2003). The role of family support, mentor support, peer behavior, and gang involvement in predicting externalizing behavior for inner-city youth. *Dissertation Abstracts International, 63*, 12-B.

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(2),* 492A.

DeBlank, G (2009). Effects of quality and quantity of the mentoring relationship on outcomes for at-risk early adolescent girls.PhD Disseratation, University of Virginia, United States -- Virginia.

DeWit, D., Lipman, E., Manzano-Munguia, M., Bisanz, J., Graham, K., Offord, D., O'Neil, E., Pepler, D. & Shaver, K.(2007) Feasibility of a randomized controlled trial for evaluating the effectiveness of the Big Brothers Big Sisters community match program at the national level. Children and Youth Services Review, 29, 383-404.

DuBois, D. L., & Silverthorn, N. (2005). Natural mentoring relationships and adolescent health: Evidence from a national study. *American Journal of Public Health, 95*, 518-524.

Frazier, C. E., Richards, P. J., & Potter, R. H. (1981). *Evaluation of Florida Project Diversion (September 1, 1980-January 31, 1981): Final Report* (1st ed). Rockville, MD: National Institute of Justice. (NCJRS Document Reproduction Service No. 81816)

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

Garate-Serafini, T., Balcazar, F. E., Keys, C. B., & Weitlauf, J. (2001). School-based mentoring for minority youth: Program components and evaluation strategies. In R. Majors (Ed.), *Educating Our Black Children: New Directions and Radical Approaches* (214-257). London: Routledge.

Gearing, B. (2008) School based mentoring and academic achievement: connecting students to the instructional program through mentoring relationships. PhD Dissertation Texas A&M, United States--Texas.

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

Goldner, L. & Mayseless, O. (2009) The quality of mentoring relationships and mentoring success. Journal of Youth and Adolescence, 38, 1339-1350Goodman, S. (1972). *Companionship therapy: Studies in structured intimacy*. San Francisco: Jossey-Bass.

Gordon, D., Iwamoto, D., Ward, N., Potts, R.,Boyd, E. (2009) Mentoring Urban Black Middle School Male Students: Implications for Academic Achievement

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 dissertation, New York University, 1979). *Dissertation Abstracts International, 41(02),* 0671B.

Green, H. (2010) The impact of an academic sports-mentoring afterschool program on academic outcomes in at-risk youth. PhD Dissertation. Drexel University--United States--Chicago, IL

Griffin, J., Holliday, R. & Frazier, E. (2009) The BRAVE (Building Resiliency and Vocational Excellence) Program: Evaluation findings for a career-oriented substance abuse and violence prevention intervention. Journal of Health Care for the Poor and Underserved, 20 (3), 798-816

Guetzloe, E. (1997). The power of positive relationships: Mentoring programs in the school and community. *Preventing School Failure, 41(3)*, 100-104.

Hanlon, T., Simon, B., O'Grady, K., Carswell, S. & Callaman, J.(2009) The effectiveness of an after-school prgram targeting urban african american youth. Education and Urban Society, 42(1), 96-111

Hart, J., O'Toole, S., Price-Sharps, J. & Shaffer, T. (2007) The risk and protective factors of violent juvenile offending: an examination of gender differences. Youth Violence and Juvenile Justice 5 (4), 367-384.Hayward, B. J., & and Tallmadge, G. K. (1995). *Strategies for keeping kids in school: Evaluation of dropout prevention and reentry projects in vocational education.* Washington, D.C.: U.S. Department of Education. (ERIC Document Reproduction Service No. ED385767)

Heard, C. A. (1990). The preliminary development of the Probation Mentor Home Program: A community-based model. *Federal Probation*, *54(4),* 51-56.

Herrera, C., Sipe, C. L., & McClanahan, W. S. (2000). *Mentoring school-age children: Relationship development in community-based and school-based programs*. Philadelphia: Public/Private Ventures.

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.

Hines, M. H. (1988). Effect of first year of matching on the self-esteem of big and little brothers. *Journal of Humanistic Education and Development, 27*, 61-69.

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)

Jent, J. & Niec, L. (2006) Mentoring youth with psychiatric disorders: the impact on child and parent functioning. Child & Family Behavior Therapy, 28(3), 43-58

Johnson, S. (2009)Therapeutic mentoring: outcomes for youth in foster care. PhD Dissertation United States—Illinois

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

Karcher, M. (2008) The cross-age mentoring program: a developmental intervention for promoting students connectedness across grade levels. Professional School Counseling, 12(2), 137-143

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.

King, K. A., Vidourek, R., Davis, B., & McClellan, W. (2002). Increasing self-esteem and school connectedness through a multidimensional mentoring program. *Journal of School Health, 72(7),* 294-299.

Klaw, E. L., Rhodes, J. E., & Fitzgerald, L. F. (2003). Natural mentors in the lives of African American adolescent mothers: Tracking relationships over time. *Journal of Youth and Adolescence, 32(3),* 223-232.

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.

Lampley, J & Johnson, K (2010) Mentoring at risk youth: improving AA in middle school students. Nonpartisan education review 6(1)

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)

Martin, A(2008) Enhancing student motivation and engagement: the effects of a multidimensional intervention. Faculty of Education and Social Work, 33, 239-269

McGreevy, S. (2007) Girl Power: The impact of an academic and social support program on the educational outcomes of urban seventh and eighth grade latinas. PhD Dissertation, State University of New York at Buffalo. United States -- New York

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.

Mecartney, C. A. (1994). *Mentoring in the Juvenile Justice System: Findings from Two Pilot Programs*. Philadelphia, PA: Public/Private Ventures.

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)

Moore, A. (2006) The impact mentoring relatinoships in an after school program has on academic achievement and behavioral growth of elementary age students. PhD Dissertation, University of Pennsylvania. United States -- Pennsylvania.

Morley, E., Rossman, S. B., Kopczynski, M., Buck, J., & Gouvis, C. (2000). *Comprehensive responses to youth at risk: Interim findings from the SafeFutures Initiative*. Rockville, MD: Juvenile Justice Clearing House. (ERIC Document Reproduction Service No. ED449294)

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

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)

Pagan, R., & Edwards-Wilson, R. (2003). A mentoring program for remedial students. *Journal of College Student Retention, 4(3),* 207-226.

Pedersen, P., Woolum, S., Gagne, B.& Coleman, M (2009)Beyond the norm: Extraordinary relationships in youth mentoring. Children and Youth Services Review, 31, 1307-1313

Postell, P. (2008)Relationships of adolescent black males and adults in the church: Influence on youths' high school completion. PhD Dissertation, University of Rochester. United States -- New York.

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)

Reglin, G. (1997). Mentoring and Tutoring Help (MATH) Program fights truancy. *Clearing House, 70(6),* 319-324.

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)

Rhoden-Trader, J. M. (1998). *Social policy and the public and private sectors: An exploratory research study on the contributions of One to One/The National Mentoring Partnership, Inc. to the mentoring movement.* (Unpublished manuscript, Union Institute and University, Cincinnati, OH).

Rhodes, J. E., Haight, W. L., & Briggs, E. C. (1999). The influence of mentoring on peer relationships of foster youth in relative and non-relative care. *Journal of Research on Adolescence, 9(2),* 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, Fort Lauderdale-Davie, FL). (ERIC Document Reproduction Service No. ED352459)

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

Rockwell, S. (1997). Mentoring through accessible, authentic opportunities. *Preventing School Failure, 41(3),* 111-114.

Rollin, S. A., Kaiser-Ulrey, C., Potts, I., & Creason, A. H. (2003). A school-based violence prevention model for at-risk eighth grade youth. *Psychology in the Schools, 40(4),* 403-416.

Rosenblum, A., Magura, S. Fong, C., Curry, P., Norwood, C. & Casella, D.(2006) Effects of peer mentoring on HIV-affected youths substance use risk and association with substance using friends. Journal of Social Service Research, 32 (2), 45-60

Roussos, S. T. (2002). How school-community partnerships contribute to widespread volunteer mentoring with youth: A study of environmental change in three communities. (Unpublished Doctoral Dissertation, University of Kansas).

Sale, E., Bellamy, N., Springer, J. & Wang, M. (2008) Quality of provider-participant relationships and enhancement of adolescent social skills. Journal of Primary Prevention, 29, 263-278

Schmidt, M., McVaugh, B. & Jacobi, J. (2007) Is mentoring throughout the fourth and fifth grades associated with improved psychosocial functioning in children? Mentoring and Tutoring, 15 (3), 263-276

Schobitz, R. P. (2004). Youth programming: An exploration of adult support processes. *Dissertation Abstracts International, 64(10-B),* 5233.

Schwartz, S., Rhodes, J., Chan, C., Herrera, C. (2011) The impact of school-based mentoring on youths with different relationshal profiles. I Developmental psychology, 47(2), 450-462

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.

Slough, N., McMahon, R. & Conduct Problems Prevention Research Group (2008). Preventing serious conduct problems in school-age youth: the fast track program. Cognitive and Behavioral Practice, 15, 3-17

Smith, D; Leve, L.; Chamberline, P. (2011)Preventing internalizing and externalizing problems in foster care as they enter middle school: impact of an intervention. Prevention Science

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.

Sterba, L. M. (2001). An evaluation of the self-esteem of at-risk middle-level youth with intervention of the Mentorship Program. *Dissertation Abstracts International*, *61(12),* 4728-A.

Struchen, W., & Porta, M. (1997). From role-modeling to mentoring for African American youth: Ingredients for successful relationships. *Preventing School Failure, 41(3),* 119-123.

Tebes, J., Feinn, R., Vanderploeg, J., Chinman, M., Shepard, J., Brabham, T., Genovese, M. & Connell, C (2007).Impact of positive youth development program in urban after-school settings on the prevention of adolescent substance use. Journal of Adolescent Health, 41, 239-247

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

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.

Welkowitz, J., & Fox, W. (2000). *A model mentor/advisor program supporting school youth with emotional and behavioral challenges and their families within rural Vermont (Final Report)*. Burlington, VT: Center on Disability and Community Inclusion, University of Vermont. (ERIC Document Reproduction Service No. ED462787).

Wilson, S.J. & Lipsey, M.W. (2007). School-based interventions for aggressive and disruptive behavior: Update of a meta-analysis. *American Journal of Preventive Medicine, 33(2S),* S130-S143.

Wunrow, J. J., & Einspruch, E. L. (2001). Promoting youth/adult partnerships: The Seven Circles Coalition in Sitka, Alaska. J*ournal of Primary Prevention, 22(2)*, 169-186.

Wyman, P., Cross, W., Brown, C., Yu, Q., Tu, X. & Eberly, S. (2010) Intervention to strengthen emotional self-regulation in children with emerging mental health problems: proximal impact on school behavior. Journal of Abnormal Child Psychology, 38, 707-720

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

8 Tables and Figures

Table 1

Categories and Variables for Meta-Analysis

Composite Category	Variables
Delinquency	
	Self-reports of delinquency
	School conduct reports
	Teacher report form (TRF [a]) or teacher BASC [b] Delinquency scales
	Arrest records
	Court records
Aggression	
	Peer nominations of aggression
	Teacher reports on the TRF of BASC
	Parent CBCL [c] or BASC reports
	Self-reports
	Behavioral Observations
Substance Use	
	Self-reports (e.g., SRD)
	Arrest records
	Court records
	Teacher reports
	Parent reports
Academic Achievement	
	School grades
	Standardized test scores (e.g., ITBS [d])
	Self-reports
	Archival graduation or withdrawal records

[a] TRF = Teacher Report Form of the Child Behavior Checklist (Achenbach, 1991)

[b] BASC = Behavioral Assessment System for Children (Reynolds & Kamphaus, 1992)

[c] CBCL = Child Behavior Checklist (Achenbach, 1991)

[d] ITBS = Iowa Test of Basic Skills (1Hieronymous, Hoover & Lindquist, 1986)

Table 2. Combinations of Search Terms Used

	Mentor	Role Model	Modeling	Interpersonal Relationship
		Delinquency		
Intervention	delinquency and mentor and intervention		delinquency and modeling and intervention	
Outreach Program Trial			delinquency and modeling and trial	
		Aggression		
Intervention				
Outreach Program		aggression and role model and outreach program		
Trial				
Psychoeducational Methods				aggression and interpersonal relationship and psychoeducational methods

Note: Combinations shown for delinquency and aggression outcomes only. Similar searches were performed for substance use and academic achievement. Derivative forms of each term were also considered.

Table 3 (in 3 sections) details of 46 studies included in Meta-Analysis

		Effect Size				Sample Size			
Citation(s) [a]	**Quality**	**Delinq.**	**Agg.**	**Acad.**	**Subs.**	**Tx**	**Con**	**Outcomes**	**Length of Follow-up**
Abbott, Meredith, Self-Kelly, & Davis (1997)	3	0.07	-0.05	0.42		22	22	Revised Problem Checklist for conduct disorder and socialized aggression; school grades	12-18 months
Aiello (1988)	3			-0.14		55	42	GPA	1 year
Anderson (1977)	3	-0.14				76	76	severity of subsequent offenses	Immediate Post-test.
Aseltine, Dupre, & Lamlein (2000)	3			0.01	.19	76	118	self-reported grades and substance use	6 month
Barnoski (2002)	3	0.22				78	78	criminal recidivism	12 months
Berger & Gold (1978)	5	0.07				46	18	Self-reported frequency of delinquency	one year
Bernstein et al (2009)	5	-0.003		-0.03		1163	1197	Self report of delinquency, school report of disciplinary action, school records of grades (math, english, science, social studies)	Post Test
Blechman, Maurice, Buecker, & Helberg (2000)	3	-0.18				45	137	Post-intake rearrest	2.5 years
Brooks (1995)	3			-0.21		23	19	GPA	Immediate Post-test.

Buman & Cain (1991)	3	0.16	0.26	137	107	High School Graduation, police arrest record.	4 years
Cavell & Hughes (2000)	5		0.02	31	29	CBCL Aggression scores.	1 year
Clarke (2009)	4	-0.24	0.80	14	11	Subject GPA, self report of "negative school behavior"	1 year
Converse & Lingugaris/Kraft (2009)	3	-1.24		16	15	Discipline Referrals	Post Test
Davidson (1976) [b] Davidson & Redner (1988) Davidson, Seidman, Rappaport, Berck, Rapp, Rhodes & Herring (1977) Ku & Blew (1977) Seidman, Rappaport & Davidson (1980)	5	1.70		25	12	Police records.	1 year

Davidson (1976) [b] Davidson & Redner (1988) Davidson, Seidman, Rappaport, Berck, Rapp, Rhodes & Herring (1977) Ku & Blew (1977) Seidman, Rappaport & Davidson (1980)	5	0.95		12	12	Police records.	1 year
Davidson & Redner (1988) Davidson, Amdur, Mitchell & Redner (1990)	5	0.60		175	85	Police Records.	2 years.
Davis (1988)	5	0.15		20	20	GPA	Immediate Post-test
Dicken, Bryson, & Kass (1977) [c]	5		0.29	20	12	Parent/teacher reports of child aggression	15 weeks
Dicken, Bryson, & Kass (1977) [c]	5		0.38	22	8	Parent/teacher reports of child aggression	15 weeks
Flaherty (1985)	5		0.00	21	21	GPA	1 year
Fo & O'Donnell (1972) Fo & O'Donnell (1975) Fo & O'Donnell (1979) O'Donnell, Lydgate, & Fo (1979)	5	-0.11		335	218	Arrests	3 years

Grant, 2010	4		0.02	13	13	Grades	Post Test
Grossman & Tierney (1998) Grossman & Rhodes (2002) Rhodes, Grossman, & Resch (2000)	5	0.08	0.11 0.18	487	472	self-reported drug use; self-reported aggressive behaviors, GPA	18 months
Hanlon, Bateman, Simon, O'Grady, & Carswell (2002)	3	0.25	0.18	214	214	Self-reported delinquency and substance use	1 year
Harmon (1995)	5		.34	43	38	Self-reported substance use.	1-20 months
Hayes (1998)	3		0.32	60	25	GPA	End of School year in which intervention occurred
Herrera, Grossman, Kauh, Feldman, McMaken & Jucovy (2007)	4	0.01	0.02	564	573	GPA	15 months
Holt, Johnson & Bry (2008)	4	-.25	-0.02	16	18	GPA & Discipline Referrals	6 months
Johnson (1997, 1999)	3		-0.03	135	171	GPA	GPA in 10th, 11th, and 12th grade (while program ongoing)
Karcher, 2008	3		-0.06	236	232	Grades	Post Test

Keating (1996) Keating, Tomishima, Foster & Alessandri (2002)	3	0.29	0.71		34	34	Self Report of Delinquency; Child Behavior Checklist (CBCL)	Immediate Post-test (after 6 months of intervention)
Kelley (1973)	5		0.28		27	22	GPA	Immediate Post-test
Kelley, Kiyak, & Blak (1979)	3	0.49			65	63	Police contacts.	Immediate Post-test
Kemple & Scott-Clayton (2004)	5		0.08		729	729	GED	48 months
Lattimore, Mihalic, Grotpeter, & Taggart (1998)	5	0.43			56	44	High School Graduation	Immediate Posttest
LoSciuto, Rajala, Townsend, & Taylor. (1996) Taylor, LoSciuto, Fox, Hilbert & Sonkowsky (1990, 1999)	3		.20		180	193	Frequency of substance use during the past 2 months.	One academic year
Maxfield, Schirm, & Rodriguez-Planas (2003)	5	0.04	-0.04	-.13	580	489	GPA, Self-reported alcohol use, self-reported criminal behavior.	Immediate to 10 months
McCord (1978, 1979)	5	-0.03			253	253	Criminal records.	30 years
Moore & Levine (1974) Moore (1987)	5	0.80			50	50	Police/court records.	1 year.

Newton (1994)	5	0.71	0.93	0.05	21	27	Violent incidents at school; Grade point average; school exclusions (suspensions)	4 months
Polit, Kahn & Stevens (1985) Quint (1991)	3			0.04	270	405	School Completion	24 months
Reyes & Jason (1991)	5			0.07	77	77	Standardized test scores.	1 academic year
Rowland (1992)	3			-0.05	42	44	School Grades	None.
Royse (1998)	5	0.35		1.43	25	21	Disciplinary infractions and GPA	10 months.
Schinke, Cole, & Poulin (2000)	3			0.62	94	94	GPA	30 months
Watson (1996)	5			0.22	69	25	GPA	6 weeks

[a] We include citations for all articles reporting results of the same studies.

[b] Based on Lipsey & Wilson, we report articles based on Davidson (1976) as two separate studies.

[c] Dickson, Bryce, & Kass (1977) reported separate analysis for males and females. Without sufficient information to combine these effects, we report them as separate outcomes for the meta-analysis.

Table 3 (second section) details of 46 studies included in Meta Analysis (cont'd)

Citation(s)	Sample Characteristics (Mentees)	Ages of Mentees	Sample Characteristics (Mentors)
Abbott, Meredith, Self-Kelly, & Davis (1997)	Boys from mother headed, single-parent households. 8-14 yrs., mean age 10 years., not diagnosed with mental or physical disabilities.	8-14	Midwestern affiliate of the Big Brothers/Big Sister of America. College educated, employed males as Big Brothers.
Aiello (1988)	underachieving students	middle school	education staff members
Anderson (1977)	Majority (97%) between 13-17 yrs., 69% male, referred by Juvenile Dept. for either criminal offenses or dependent-incorrigible (runaway, truant). pg. 49 - Age: 14.13, Severity of Original offense: 3.71 (scale 1-5), 51% male, 48 % female.	13-17	Volunteers recruited through friends of court, speaking engagements, and universities, screened through multiple interviews, and identified as having sincere interest in helping someone have a productive life. White collar professionals, between 26-35 yrs.
Aseltine, Dupre, & Lamlein (2000)	Low income 6th grade students living in large urban setting.	12 to 13	Adult mentors over age 50.
Barnoski (2002)	Juveniles from juvenile confinement. Minimum of 5-6 months in Juvenile confinement remaining, non-sex offenders.	Under 18	Trusted adult volunteer to help youth set educational and vocational goals, and live drug-free, crime-free life. One year commitment; recruited through posters, service groups, internet, and from pastors. Mostly female.
Berger & Gold (1978)	Juvenile court-selected probationers.	Under 18	community volunteers
Blechman, Maurice,	Minors charged with nonviolent misdemeanors or first felonies	8.85 - 18.33	Adult volunteers

Buecker, & Helberg (2000)	Participant gender was 71.8% male (n = 176). Ethnicity was 76.7% white (n = 188), 17.1% Latino (n = 42), and 6.1% black, Asian, Native American, and multi-ethnic (n = 15).		
Bernstein et al (2009)	4th-8th graders referred for school failure, low self esteem or lack of role models. 82% recieving free/ reduced lunch. 41% african american, 23% white, 29% hispanic. 57% female	4th-8th grade	Volunteers from community. Range of recruitment strategies. 72% female and 66% white. Majority were college aged.
Brooks (1995)	High school students nominated by teachers based on academic performance and extracurricular activities. From economically disadvantaged schools, primarily African American (89%) and female (86%).	15 - 18 years old	College student volunteers for mentoring program (81% female, 19% male, aged 19-42, 74% African-American, 26% White, min. 2.5 GPA)
Buman & Cain (1991)	Summer Youth Employment Program (SYEP) participants (low family income) who had automatically assigned Business Partners during the summer of 1986, and whose workplace included access to phones and were employed for more than 6 weeks. (Control group randomly selected from remaining files, and those who did not have Business Partners in subsequent years). 14-21 yrs old. Approx. 70% Black, 9% Asian, <1% Hispanic, 11% Native American, 8% White; nearly equal male/female. Ave. age 16, majority (59%) from mother-only households.	14-21	Volunteer mentors recruited by private sector companies participating in program to employ (at minimum wage) youth whose household incomes are below poverty level in Minneapolis.
Cavell & Hughes (2000)	> 84th percentile Aggressive Behavior scale of the Teacher Report Form. Primarily African American (48%) and White (37%); and Male (77%).	Grade 2-3	College undergraduate students fulfilling course requirements.
Davidson (1976) Davidson & Redner	Local youth contacted by juvenile bureaus and considered in jeopardy of juvenile court referral. Mostly white (76%) and males (76%).	Mean age of 14.1 years.	College students matched on mutual interests, sex, and race.

(1988) Davidson, Seidman, Rappaport, Berck, Rapp, Rhodes & Herring (1977) Ku & Blew (1977) Seidman, Rappaport & Davidson (1980)			
Clarke 2009	Mentees: N= 18, girls and boys. 9th graders identified by teachers for behavior problems and risk of failing/ dropping out school	9th Grade	Mentors: Teachers and other school staff volunteers from study sites. (12 teachers & 1 guidance counselor) 50% male & 50% AA.
Converse & Lingugaris/Kraft (2009)	Mentored group 56% white, 44% Hispanic. Control group 40% white, 60% Hispanic. 80% male		Mentoring provided by 13 faculty/ staff from school - all white, 11/13 female.
Davidson (1976) Davidson & Redner (1988) Davidson, Seidman, Rappaport, Berck, Rapp, Rhodes & Herring (1977) Ku & Blew (1977) Seidman, Rappaport & Davidson (1980)	Youth from low income families with prior arrests. Mostly male (92%); white (58%) and African American (42%).	Mean age-14.5 yrs.	College students matched on mutual interests, sex, and race.

Davidson, Amdur, Mitchell & Redner (1990) Davidson & Redner (1988)	Juveniles referred from local juvenile court. Mostly males (84%) and White (77%)	Mean age of 14 years	College students and some community volunteers.
Davis (1988)	Students repeating 9th grade. Mostly males (60%); African American (46%) and White (54%)	Mean age of 15.6 years	Volunteer teachers and school staff.
Dicken, Bryson, & Kass (1977)	Families of elementary school age *boys* from low-income families. Most families were headed by single mothers. All were Caucasian.	6 - 13 years old	College students ; must be juniors or seniors; same sex as child; demonstrate motivation by attending orientation and supervisory sessions.
Dicken, Bryson, & Kass (1977)	Families of elementary school age *girls* from low-income families. Most families were headed by single mothers. All were Caucasian.	6 - 13 years old	College students ; must be juniors or seniors; same sex as child; demonstrate motivation by attending orientation and supervisory sessions.
Flaherty (1985)	Random sample of basic academic level (math and science) freshman students. 71% white, 28.5% black, Asian, or other. 52% low socioeconomic class, 19% inner city, 16.5% middle class, 7.5% mid-high class, 5% high socioeconomic class.	14-15	Members of the teaching staff at a high school.
Fo & O'Donnell (1972) Fo & O'Donnell (1975) Fo & O'Donnell	Youth referred based on behavior and academic problems (truancy, poor academic achievement, classroom disruption, curfew violation, fighting). Ave. age 14 (7th, 8th grade). Hawaiian, Filipino, Japanese, Chinese, and Caucasian.	11-17	Adult residents of the community recruited through newspaper ads. Aged 17-65, both sexes, diverse group ethnically and occupationally. Education range from 4th grade-master's degrees (median achievement 12th grade).

(1979) O'Donnell, Lydgate, & Fo (1979)			
Grant, 2010	8-6th grade African American boys who were at risk for school failure (GPA lower than 2.0) who were nominated by teachers and principles.	6th-8th grade	Spirituality and knowledge of African American culture were integrated into group meetings and individual mentoring.
Grossman & Tierney (1998) Grossman & Rhodes (2002) Rhodes, Grossman, & Resch (2000)	Majority of boys (62%) and Minority (not specified, 57%)	10-16 years old	well educated young professionals, incomes > 40,000
Hanlon, Bateman, Simon, O'Grady, & Carswell (2002)	Inner-city youth referred as at risk for developing a deviant lifestyle and met one or more criteria: alcohol or drugs, history of delinquency or other deviant behavior, expulsion from school. 97.4% black, 2.6% white; 59% male, . 50% referred by family, 26% by school, 17% from community agency, 6% by juvenile justice system. 2/3 had been arrested before.	9-17 yo	Mentoring positions staffed by representative role models from community (young African-American college students) who were available during group mentoring sessions 4-5 days/wk. after school with 20 kids/session. Staff/child ratios 1:8 (never less than 1:10).
Harmon (1995)	Pregnant and parenting teens and young adults of Harford County. 98% female; 48% white, 50% black, 2% other; 80% unemployed; 42% pregnant	14-21 yo	Community volunteers whose work and family lives resemble participant's goals for future. Prosocial, positive role model volunteers
Hayes (1998)	Students identified by their counselors prior to entering 9th grade as being "at risk" of dropout; ave. family income in low average range, 25% students on free or reduced lunch, 11% absentee rate) More males identified as at	grades 9-12	Volunteer staff members from the high school, including teachers and support personnel. Only requirement was to mentor student for years student remained in school.

	risk than females.		
Herrera, Grossman, Kauh, Feldman, McMaken & Jucovy (2007)	40 students chosen from 97 9th graders who were completing a universal prevention program "Peer Group connection" for high school transistion. At risk for academic failure.	9th graders	Teacher/ staff at school who could committ to 4 hours a week. 10 mentors for 20 students
Johnson (1997, 1999)	At-risk youth based on recommendations from jr. high or high school teachers/counselors. Half male, half female, 75% black, middle-achieving students (B-C range GPAs), qualify for free or reduced-price lunch program	grades 9-12	Mentors recruited through personal contacts, program presentations, TV and radio solicitations. Matched on gender (but not race). Initial training session and monthly check-ins with staff. Most over 45 yrs. Old, predominantly white, with older children who no longer live at home. Live and work in city. 1/3 of mentors personally made $6,000 contribution toward student's financial support; 1/3 previously involved in another mentoring program.
Karcher, 2008	5th - 8th graders in large southweestern city . Majority from low income families. Majority Mexican American or Hispanic/ anglo biracial	5th - 8th grade	School based mentors were 54% latino, 3% caucasian, 5% AA and 6% other. 70% college students. 43% spoke Spanish. 73% female.
Keating (1996) Keating, Tomishima, Foster & Alessandri (2002)	Youth deemed at-risk for juvenile delinquency or mental illness (but not involved in serious delinquent behavior). 65% male, 35% female; 32% white, 24% black, 37% Latino, 3% Asian, 3% other.	10-17 yo	Adults who live in surrounding communities and interested in helping troubled youth. Must be over 18 yrs., screened for commitment to program and appropriateness for involvement with at-risk youth. Mentors matched as close as possible on gender, ethnicity, age, geographical location, and common interests.
Kelley (1973)	Boys referred from court intake - deemed not serious enough for court hearing, but needing intervention. Mean age 14 yrs., 59% referred as 1st	10-16 yo	Undergraduate males from 2 urban colleges, all enrolled in psychology courses and volunteering as a course requirement.

	offense, equal number black/white.		Mean age 27.5 years.
Kelley, Kiyak, & Blak (1979)	Youth in juvenile court diversion progam. No more than 3 "unofficial" police contacts; no formal adjudication hearings at juvenile court; voluntary admission to the program; no extreme disabilities; ages 10-17 yrs. Mean age 14.5 yrs. 78% black, 22% white. Equal male, female.	10-17 yo	Students from 2 urban colleges as volunteer counselors (1/2 4 yr. college, 1/2 community college), all enrolled in psychology courses and volunteering as a course requirement.
Kemple & Scott-Clayton (2004)	High school students in a large urban school district.	14-22	Employer Partners
Lattimore, Mihalic, Grotpeter, & Taggart (1998)	Youth from low income families receiving public assistance in 5 industrialized areas. Youth enter program as freshman in high school, and program continues through 4 years of high school.	14-20	Mentor is "Coordinator"; prosocial adult who acts as surrogate parent, role model, advisor, and disciplinarian to "Associate" (youth)
LoSciuto, L., A.K. Rajala, T.N. Townsend, and A.S. Taylor. (1996)	6th graders from low income communities. Primarily African American.	11-12	Volunteers ranging in age from 60-85. Primarily African American from low income communities. Asked to make 1 year commitment.
30Maxfield, Schirm, & Rodriguez-Planas (2003)	Youth entering 9th grade at a high school with dropout rates > 40%. Youth were not repeating 9th grade, did not have disabilities that would interfere with participation, and had GPA < 67th percentile.	14-18	Mentors were case managers with caseloads of 15-25.
Moore & Levine (1977) Moore (1987)	Selected by probation officer to be at high risk for re-offending. All were white males.	16-22 yo	Citizen volunteers matched to sex, ethnicity, education/vocation, and interests.
Newton (1994)	Middle school students selected on basis of school failure and history of	grades 7-8	College students; primarily men (83%) and African American

	violent behavior. Primarily male (73%) and African American (73%).		(67%).
Polit, Kahn & Stevens (1985) Quint (1991)	Primarily low-income African American and Latino women who we pregnant or parenting at the time of study enrollment.	14-17	Volunteer women from low-income neighborhoods. Range in age from 20s to 70s, most with high school diploma, but not working. Matched to between 1 and 5 teens.
Reyes & Jason (1991)	Ninth grade students from a large urban school with a high (60%) dropout rate. Primarily Hispanic.	9th grade	Homeroom teachers - trained to provide guidance and counseling.
Rowland (1992)	Identified as high-risk of dropping out of school before graduation.	grades 1-5	Area business men and women, community leaders, retirees, and civic members.
Royse (1998)	African American teenagers, ages 14-16 from female-headed household and less than grade equivalency in reading, math, and science. Live in household with income at or below 125 % federal poverty guidelines.	14-16 years	African American male community volunteers. Most were college graduates in their 30s.
Schinke, Cole, & Poulin (2000)	40% female, ave. age 12.3 yrs., 63% black, 19% Hispanic, 13% white, 5% Asian and other.	12.3 avg	Boys and Girls Club of American staff assisted by parents and other volunteers
Watson (1996)	Hispanic middle school and high school students identified as "at-risk" at least one of the characteristics: 1. retained at least one grade, 2. 2 or more yrs below grade level in standardized tests, 3.failed at least 2 courses, 4. failed at least one section of the statewide standardized test.	middle/high school	Senior citizen and college student mentors recruited throughout community.
McCord (1978, 1979)	Boys from densely-populated urban industrial areas identified by schools, welfare agencies, churches and police as "difficult" or "average", given physical exams, and then matched in pairs on age, delinquency-prone histories, family background, and home environments (coin toss determined	5-13 yrs. Original study; 35-44 for follow up	Social workers who tried to form a close personal relationship with boy and help with the boy and his family in variety of ways. Counselors not allowed to have contact with criminal justice agencies or with boys in control group.

group).

Table 3 (third section) details of 46 studies included in Meta Analysis (cont'd)

Citation	Description of Mentoring	Additional Interventions
Abbott, Meredith, Self-Kelly, & Davis (1997)	Big Brothers/Big Sisters of America. Adult companion program; weekly companionship between boy and adult male for 12-18 months for a visit and/or activity. Big Brother to serve as positive role model to child in vocational, psychological, and social ways.	None
Aiello (1988)	take part in a series of structured and unstructured activities throughout the years. Bimonthly meetings between mentors and mentees.	None
Anderson (1977)	One to One program - volunteers spent at least 2 hrs. per week and were looked at role models, friend, or assistant	Family Crisis Intervention: Serving "most in need" families (children refusing to go home with families) with therapy sessions; at publication, data not yet available.
Aseltine, Dupre, & Lamlein (2000)	Mentors spend at least 2 hours/week in one on one contact with youth. Activities include tutoring, community service, recreational activities, and assistance with school projects.	Community Service (youth spend 1-2 hrs./wk.), Social Competence Training (26 weekly lessons include stress management, self-esteem, etc)., Family Activities (monthly weekend events for youth, their families, and their mentors).
Barnoski (2002)	Meet monthly during last 5-6 months of youth confinement in Juvenile facility	None
Berger & Gold (1978)	One on one similar to Big Brothers/Big Sisters	Some (number not specified) chose to participate in group counseling or tutoring.
Bernstein et al (2009)	103 schools participating in federally funded examination of the	None

	effectiveness of school based mentoring. The programs focused on academic goals, self esteem, relationship building, and giving advice.	
Blechman, Maurice, Buecker, & Helberg (2000)	Adult volunteers who spent 2 hours a week for approximately 21 weeks with proteges Mentors attended a training program	None. all participants received Juvenile Diversion program. Study compared JD to JD+Mentoring
Brooks (1995)	take part in a series of structured and unstructured activities throughout the years. Bimonthly meetings between mentors and mentees.	None
Buman & Cain (1991)	Volunteer mentors commit to meet Youth Partners at youths' worksites, contact them by phone once/week to discuss work issues, accompany them to work sponsored events.	none
Cavell & Hughes (2000)	"Therapeutic" mentors received 18 hours of training. Mentor visits were at least 1 hour per week outside of school hours for 16 months of intervention. Goal of providing accurate understanding, emotional acceptance, and firm limits on antisocial behaviors. Engaged in interactive activities.	Treatment group received therapeutic mentors (i.e., highly trained and supervised), teacher consultation, parent consultation, and problem-solving skills training. Control group received "standard" (i.e., untrained and unsupervised) mentoring.
Clarke, 2009	"Achievement Mentoring" - took palce during the second semseter of the ninth grade. Adaptation of "Behavioral Monitoring and reinforcement program). Mentors spoke with teachers, meet with mentee for 20 minutes and follow up on achievement and goals.	
Converse & Lingugaris/Kraft (2009)	Mentoring occurred over 18 weeks, for an average of 15 meetings. Involved relationship building, support and academic	

Davidson (1976) Davidson & Redner (1988) Davidson, Seidman, Rappaport, Berck, Rapp, Rhodes & Herring (1977) Ku & Blew (1977) Seidman, Rappaport & Davidson (1980)	Relationship building, behavioral contracting, and child advocacy.	Community Advocacy (targeting of community resources).
Davidson (1976) Davidson & Redner (1988) Davidson, Seidman, Rappaport, Berck, Rapp, Rhodes & Herring (1977) Ku & Blew (1977) Seidman, Rappaport & Davidson (1980)	Relationship building and child advocacy.	Community Advocacy (targeting of community resources).
Davidson, Amdur, Mitchell & Redner (1990)	Relationship building, behavioral contracting, and child advocacy.	None

Davidson & Redner (1988)		
Davis (1988)	Relationship building, support, attendance and academic monitoring.	None
Dicken, Bryson, & Kass (1977)	companionship program, 2 visits and 6 hrs. of contact per week during an academic semester in a variety of settings.	none
Dicken, Bryson, & Kass (1977)	companionship program, 2 visits and 6 hrs. of contact per week during an academic semester in a variety of settings.	none
Flaherty (1985)	Members of the teaching staff served as advocating adults for mentees.	None
Fo & O'Donnell (1972) Fo & O'Donnell (1975) Fo & O'Donnell (1979) O'Donnell, Lydgate, & Fo (1979)	Adult buddies attempted to influence youth through their relationship and contingent use of social and material reinforcement. Buddies paid $144/month by earning points for training, contact, and documentation.	contingent material reinforcement for target behaviors
Grant, 2010	Christian African American community based mentoring. Offers peer group involvement, skill development, knowledge of African American culture and mentoring.	Spirituality and knowledge of African American culture were integrated into group meetings and indivdual mentoring.
Grossman & Tierney (1998) Grossman & Rhodes (2002) Rhodes, Grossman, &	Big Brother Big Sisters program (BBBS), 3-4 hr. meetings 2-4 times per month for at least one year.	none

Resch (2000)		
Hanlon, Bateman, Simon, O'Grady, & Carswell (2002)	Group mentoring session 4-5 days/week after school. Homework help, regularly scheduled activities and presentations, holiday parties, field trips.	All subjects received individual counseling. Counselors in the experimental clinic were trained in specific case management strategies, were provided support in identifying community resources. Counselors also led informal discussions about parenting and led program-sponsored parent/child social events. Subjects in the experimental condition also received remedial education.
Harmon (1995)	Goal to provide opportunity for youth to bond with prosocial others, increase self-esteem, life management, and employability skills, and decrease favorable attitudes toward drug use.	Drug education, monthly career and/or personal development workshops, "Bright Futures" curriculum for high risk youth (using worksheets, discussion, role play for sessions ranging from self-esteem to drug abuse education), Peer Leadership Training (after 80% completion of Bright Futures program; includes weekend retreat).
Hayes (1998)	Staff met 4 times for 1 hour during 1st year to receive training in at-risk student behavior. Mentors to spend as much time with mentee as they feel comfortable. Mentors provided support and guidance to their student mentees by placing emphasis on interpersonal relationships, problem solving techniques, communication skills, positive behavior, study skills.	None
Herrera, Grossman, Kauh, Feldman,	School based mentoring program through the use of Big Brothers Big Sisters mentoring program. Meet once a week at school with mentor	Tutoring with mentor

McMaken & Jucovy (2007)	during or after school from 30-60 minutes. Completed social and academic activities as pairs and with groups of mentor matches.	
Holt, Bry & Johnson (2008)	"Achievement Mentoring" - took place during the second semester of the ninth grade. Adaptation of "Behavioral Monitoring and reinforcement program). Mentors spoke with teachers, meet with mentee for 20 minutes and follow up on achievement and goals.	
Johnson (1997, 1999)	Mentors meet with mentees at least once monthly, with phone calls in between meeting times. Provide assistance in college and financial aid application process, attend SAS outings, monitor student's grades, and report on relationship's progress with SAS program staff.	None
Karcher (2008)	8 meetings with school-based mentor. Part of larger multii-component school based intervention. Mentors were 54% latino, 3% caucasian, 5% AA and 6% other. 70% college students. 43% spoke spanish. 73% female.	Both experimental and comparison group received supportive services through community agency including educational enhancement activities, guidance, enrichment activities or tutoring.
Keating (1996) Keating, Tomishima, Foster & Alessandri (2002)	Youth and adults spend a minimum of 3 hrs. in activities such as going to sporting event, the movies, or a park.	Life skills training - a monthly seminar conducted by local professionals on topics such as: child abuse prevention, drug and alcohol abuse, cross cultural awareness, health, nutrition, and school problems.
Kelley (1973)	Ultimate goal for each student counselor was to establish with his juvenile companion, "corrective counseling relationship." 1:1 mentors, 3-8 months (ave. 5.6 months), 4 times/month, less than 3 hrs. each meeting.	None

Kelley, Kiyak, & Blak (1979)	Meetings weekly for a minimum of 4 hrs.	None
Kemple & Scott-Clayton (2004)	Interpersonal support.	Implemented Career Academies - a school-within-a-school organization and that also provided i
Lattimore, Mihalic, Grotpeter, & Taggart (1998)	"Coordinator", or mentor, coordinates the program for youth partner. 250 hrs. educational activities (computer-assisted instruction, peer tutoring); 250 hrs. development activities (cultural activities, acquiring life/family skills, college and/or occupational training); 250 hrs. service activities (community service projects, helping with public events, work as volunteer for various agencies).	Education activities (e.g., peer tutoring, computer-assisted instruction), development activities (e.g., planning for college, job preparation), service activities (e.g., community service, volunteering). Financial Incentives.
LoSciuto, L., A.K. Rajala, T.N. Townsend, and A.S. Taylor. (1996)	Spent a minimum of 4 hours together each week, engaging in a variety of activities (.e.g. helping with homework, attending class field trips, attending cultural/sporting events).	None. Treatment condition consisted of mentoring plus community service, classroom-based life skills curriculum, and parent workshops. Control group received community service, classroom-based life skills curriculum, and parent workshop interventions without mentoring.
30Maxfield, Schirm, & Rodriguez-Planas (2003)	No description of specific mentoring activities other than mentoring being a component of the case management. Noted that case managers developed "deep personal relationships" with 40 - 60 percent of students at some sites.	Case management, target of 250 hours in each of 3 service components - education, developmental activities, and community service. Financial incentives.
Moore & Levine (1977) Moore (1987)	Weekly meetings between "citizen counselors" and subjects.	None - probation programs for all individuals. Citizen counseling only to treatment group.
Newton (1994)	Each mentor met weekly with 1-2 students during 1 semester. Provided	None

	academic assistance, worked with teachers to establish behavioral goals, and served as positive role models.	
Polit, Kahn & Stevens (1985) Quint (1991)	Served as confidantes, escorted to appointments, recreational events, made reminder calls, and acted as paraprofessional case managers.	Informational workshops, links to services, and individual counseling.
Reyes & Jason (1991)	Guidance and counseling by homeroom teachers.	Redesign of school day to keep students in homeroom together (3 core classes). Feedback to parents every 5 weeks.
Rowland (1992)	Mentors met with high-risk students for min. of 1 hr./wk. for school year.	None
Royse (1998)	No details on content of mentoring. Also included monthly group outings.	None
Schinke, Cole, & Poulin (2000)	Discussions with adults.	Weekly structured activities of the educational enhancement program.
Watson (1996)	Four distinct mentoring treatments: (1) Mentor called student 2x/wk. (2) Student instructed to call mentor 2x/wk. (3) Mentor met with group of 5 students 2x/wk. (4) Mentor met with student 2x/wk.	none
McCord (1978, 1979)	5 year treatment; counselors assigned to each family visited ave. twice a month.	For treatment group, 1/3 focused on family problems, 1/2 boys tutored in academic subjects, 1/2 received medical or psychiatric attention, 1/4 sent to summer camps, most brought into Boy Scouts, YMCA, or similar community programs.

Table 4

Standardized Mean Difference Effect Sizes and Homogeneity Statistics from Random Effects Mentoring Meta-Analyses

Model	SMD	95%CI	Z	τ^2	I^2	**H**	**Q**	*df*
Delinquency (k = 25 studies)	0.21	0.17 – 0.25	9.84**	0.01	99.3%	11.72	3297.64**	24
Aggression (k = 7 studies)	0.29	-0.04 – 0.62	1.71	0.18	95.4%	4.66	130.35**	6
Drug Use (k = 6 studies)	0.16	-0.00 – 0.32	1.93	0.04	99.9%	27.01	3647.33**	5
Academic Achievement (k = 25 studies)	0.11	0.07 – 0.15	5.86**	0.01	60.0%	1.58	59.95**	24
Overall Effects (k = 46 studies)	0.18	0.15 - 0.21	10.80**	0.01	99.2%	11.50	5948.72**	45

Note: * $p < .05$, ** $p < .01$

Table 5

Differences in Mean Effect Sizes by Study Design						
	Quasi-experimental Designs		Randomized Controlled Trials		*Meta-Regression*	
	# Studies	*SMD*	# Studies	*SMD*	*B*	*SE*
Delinquency	11	0.20	14	0.42	.19	.16
Aggression	3	0.14	4	0.41	.26	.29
Drug Use	3	0.19	3	0.13	-.07	.11
Academic Achievement	15	0.14	10	0.22	.19	.16

Table 6

Moderation of Mentoring Effects (Random Effects Models)										
	Level of Moderator								Meta-Regression	
	Absent				Present					
Moderator	***k***	***SMD***	***L***	***U***	***k***	***SMD***	***L***	***U***	***B***	***SE***
Mentee Selection										
Individual Risk	22	0.20	0.05	0.35	16	0.23	0.11	0.35	0.03	0.09
Environmental Risk	28	0.20	0.09	0.31	8	0.23	-0.06	0.51	0.03	0.12
Other Interventions	23	0.20	0.06	0.34	23	0.31	0.13	0.49	0.07	0.10
Motivations of Mentors										
Civic Duty	11	0.24	0.00	0.47	21	0.22	0.09	0.35	0.02	0.11
Professional Development	20	0.16	0.05	0.27	16	0.42	0.16	0.68	0.21*	0.11
Quality and Fidelity Checks										
Quality Check	14	0.20	0.06	0.35	20	0.21	0.05	0.38	-0.00	0.10
Fidelity Check	27	0.20	0.09	0.30	6	0.29	-0.15	0.73	0.05	0.14
Key Processes										
Modeling/Identification	28	0.24	0.08	0.40	11	0.32	0.08	0.56	0.06	0.12
Emotional Support	12	0.11	0.00	0.23	27	0.32	0.14	0.50	0.22*	0.12
Teaching	11	0.12	-0.01	0.24	30	0.29	0.15	0.44	0.15	0.10
Advocacy	32	0.13	-0.05	0.31	10	0.39	0.06	0.72	0.17*	0.09

Notes: * $p < .05$, one-tailed

Random effects models of standardized mean differences (SMD) are the sources of the significance tests for the SMDs within levels of each moderator. The meta-regression models are mixed effects models using full maximum likelihood estimation. *k* = number of studies, *SMD* = standardized mean difference, *L* = lower limit of the 95% confidence interval for the *SMD, U* = upper limit of the 95% confidence interval for the *SMD*

Table 7: Citations for Excluded Studies

Excluded Studies (N = 118)
Abcug (1991)
Ahrens, Richardson ,Lorazno, DuBois (2007)
Baldwyn Separate School District, MS. (1982).
Banta & Lawson (1980)
Barron-McKeagney, Woody, & D'Souza (2001)
Beier, Rosenfeld, Spitalny, Zansky, & Bontempo (2000)
Bellamy, Springer, Sale, & Espiritu, (2004)
Bernstein, Rappaport, Olsho, Hunt & Levin (2009)
Bilbrew (2009)
Blakely, Menon, & Jones (1995)
Blinn-Pike, Kuschel, McDaniel, Mingus, & Mutti (1998).
Bracy (2008)
Bruce & Mueller (1994)
Campos,Phinney, Perez-Brena, Kim, Ornelas, Nemanim, Padilla, Mihecoby & Ramirez (2009)
Carrington, Tymms, & Merrell (2008)
Cave & Quint (1990)
Cheng, Haynie, Brenner, Wright, Cung & Simons-Morton (2008)
Ching, Yeh, Siu, Wu & Okubo (2009)
Clarke (2009)
Colley (2003)
Colley (2003)
Colson, Godsey, Mayfield, Nash, & Borman (1978)
Conduct problems prevention research group (2007)
Conduct problems prevention research group (2010)
Cummings (2010)
Dance (2001).
Dappen & Isernhagen (2002).

Davis, & Haney (2003)
Davison (1994)
De Blank (2009)
De Wit, Lipma, Manzano-Munguia, Bisanz, Graham, Offord, O'Neill, Pepler & Shaver (2007)
DuBois & Silverthorn (2005).
Elledge, Cavell, Ogle & Newgent (2010)
Frazier, Richards, & Potter (1981) – 2 studies
Galvin (1989)
Garate-Serafini, Balcazar, Keys, & Weitlauf (2001)
Gearing (2008)
George (1986)
Goldner & Mayseless (2009)
Gordon, Iwamoto, Ward, Potts & Boyd (2009)
Goodman (1972)
Graber (1985)
Grant (2010)
Green (1979)
Green (2010)
Guetzloe (1997)
Hanlon, Simon, O'Grady, Carswell & Callaman (2009)
Hart, O'Toole, Price-Sharps & Shaffer (2007)
Hayward & Tallmadge (1995)
Heard (1990)
Hernandez (2009)
Herrera, Sipe, & McClanahan (2000)
Herrera, Grossman, Kauh & McMaken (2011)
Hill (1972)
Hines (1988)
Howitt, Moore, & Gaulier (1998)
Huisman (1992)

Jent & Niec (2006)

Johnson (2009)

Johnson, Holt & Powell (2008)

Joseph (1992)

Karcher (2008)

Keenan (1992)

King, Vidourek, Davis, & McClellan (2002)

Klaw, Rhodes, & Fitzgerald (2003)

Komro, Flay, & Biglan (2011)

Lakes (1997)

Lamb (2010)

Lampley & Johnson (2006)

Laughrey (1990)

Lee, Plionis, & Luppino (1989)

Martin (2008)

McGreevy (2007)

McPartland & Nettles (1991)

Mecartney (1994)

Mertens (1988)

Mitchell, & Casto (1988)

Morley, Rossman, Kopczynski, Buck, & Gouvis (2000)

Nelson, & Valliant (1993)

New York City Board of Education (1986)

Pace (2010)

Pagan & Edwards-Wilson (2003)

Pedersen, Woolum, Gagne & Coleman (2009)

Postell (2008)

Powers & McConner (1997)

Powers, Sowers, & Stevens (1995)

Reglin (1997)

Reller (1987)

Rhoden-Trader (1998)

Rhodes, Haight, & Briggs (1999)

Rippner (1992)

Roberts, & Cotton (1994)

Rockwell (1997)

Rollin, Kaiser-Ulrey, Potts, & Creason (2003) – 2 studies

Rosenblum, Magura, Fong, Curry, Norwood & Casella (2006)

Roussos (2002)

Rulison (2010)

Sale, Bellamy, Springer, Wang (2008)

Schmidt, McVaugh & Jacobi (2007)

Schwartz, Rhodes, Chan & Herrera (2011)

Schobitz (2004)

Seidle (1982)

Slicker & Palmer (1993)

Slough, McMahon & Conduct Problems Prevention Group (2008)

Smith (1990).

Smith, Leve & Chamberline (2011)

Stanwyck & Anson (1989)

Sterba (2001)

Struchen & Porta (1997)

Tebes, Feinn, Vanderploeg, Chinman, Shepard, Brabham, Genovese, Connell (2007)

Tierney, Grossman, & Resch (1995)

Turner & Scherman (1996)

Valenzuela-Smith (1984)

Welkowitz & Fox (2000)

Wunrow & Einspruch (2001)

Wyman, Cross, Brown, Yu, Tu & Eberly (2010)

Wyatt (2009)

Zand, Thomson, Cervantes, Espiritu, Klagholz, LaBlanc & Taylor (2009)

9 Figure Captions

Figures 1-4

Forest plots of meta-analysis of the effects of mentoring interventions for each outcome.
Figure 1 reports studies measuring outcomes related to delinquent involvement.

Figure 2 reports effects related to academic achievement.

Figure 3 reports effects on aggression or externalizing behaviors.

Figure 4 reports effects on illegal drug use. The size of the center square shows the weight assigned to the study and the width of the error bars shows the 95% confidence interval for the effect size of each study.

Figures 5-6

Plots of average overall standardized mean difference (SMD) effect sizes and 95% confidence intervals by levels of moderating variables.

Figure 5 graphs moderation of overall effects by two possible motivations of mentors, civic duty and professional development.

Figure 6 graphs the overall effect estimates by the presence or absence of key processes in the mentoring intervention, including emotional support, promotion of modeling or identification with the mentor, and teaching.

Figure 1

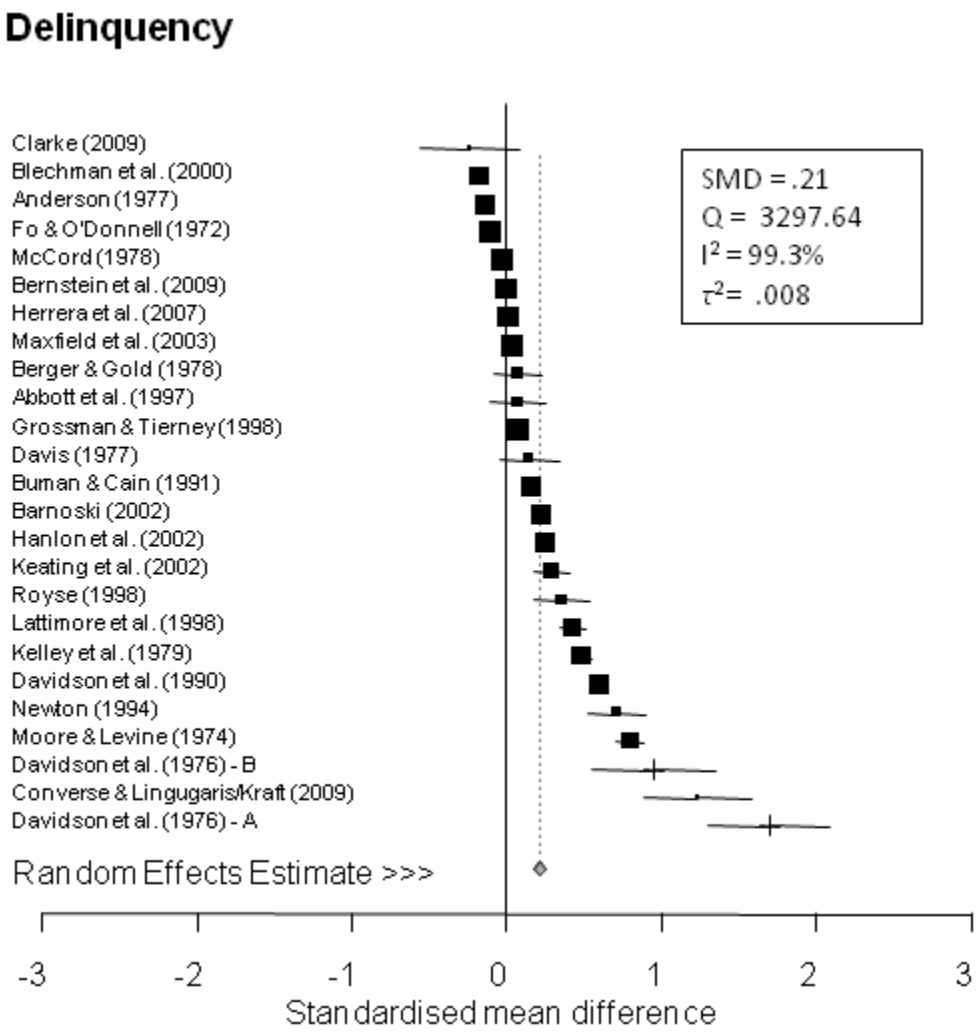


Figure 2

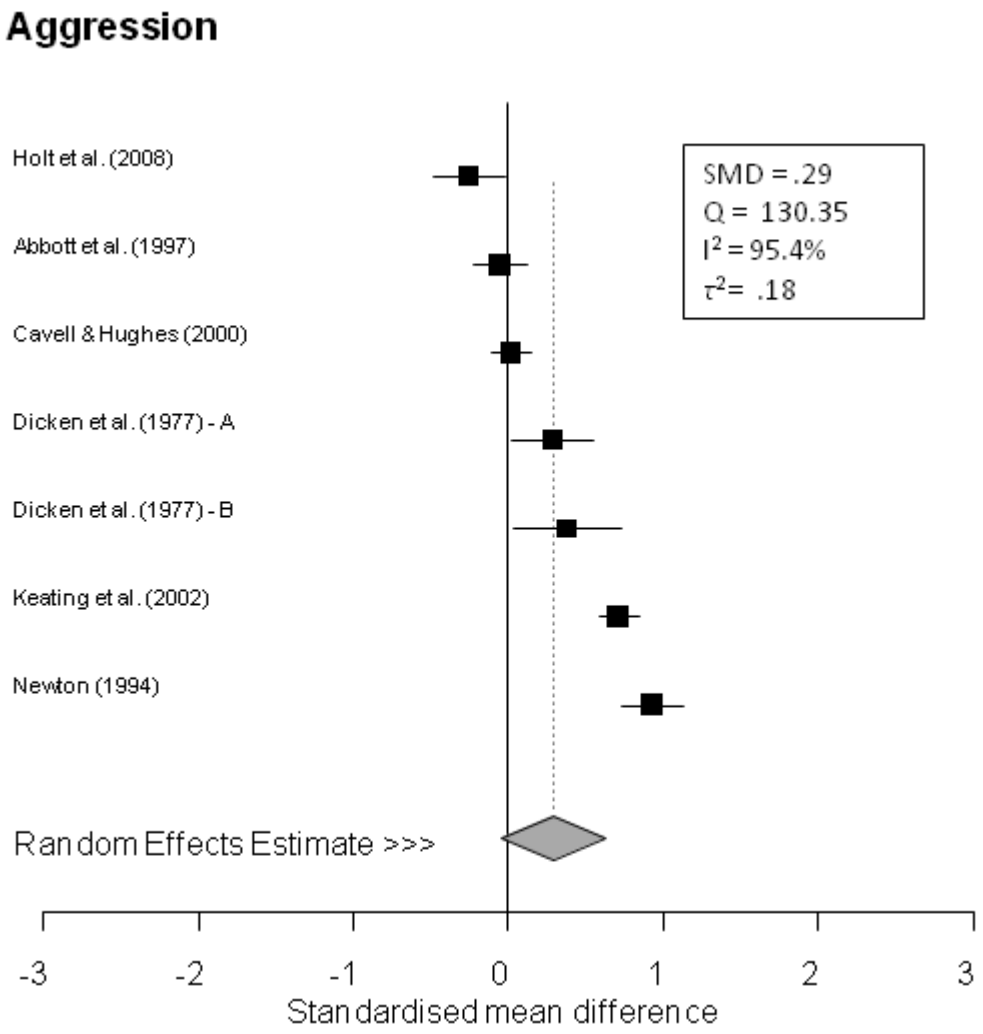


Figure 3

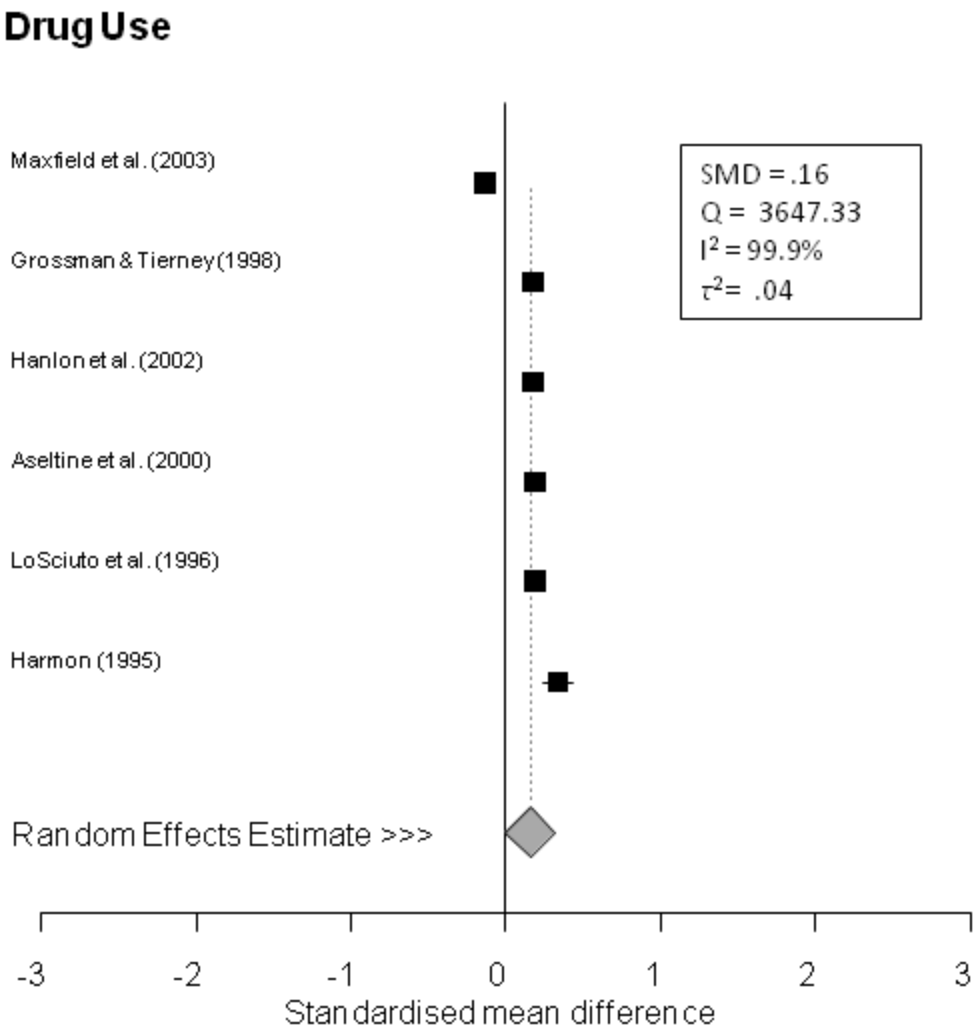


Figure 4

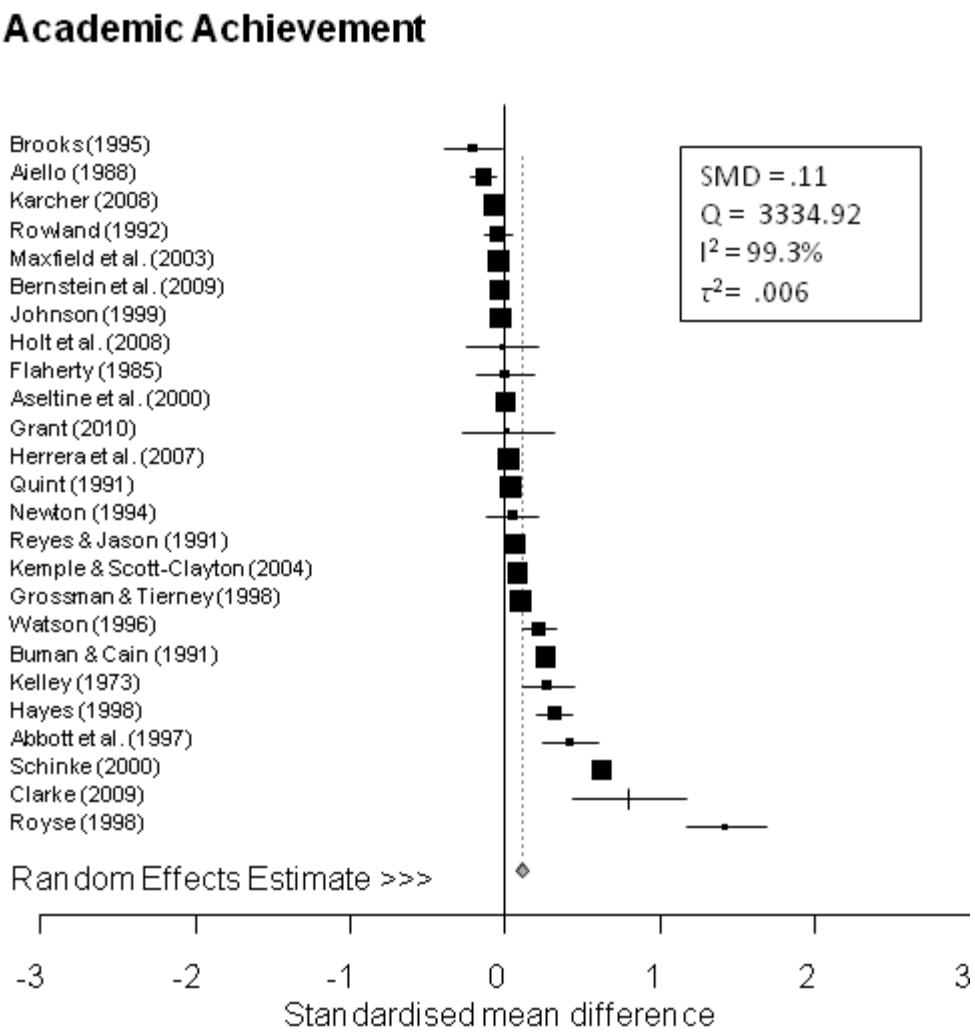


Figure 5

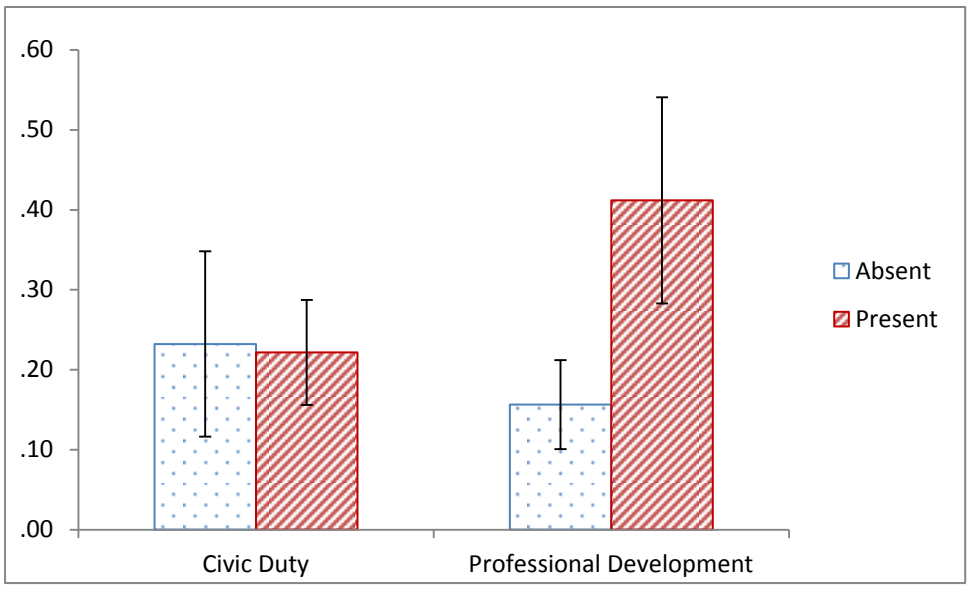
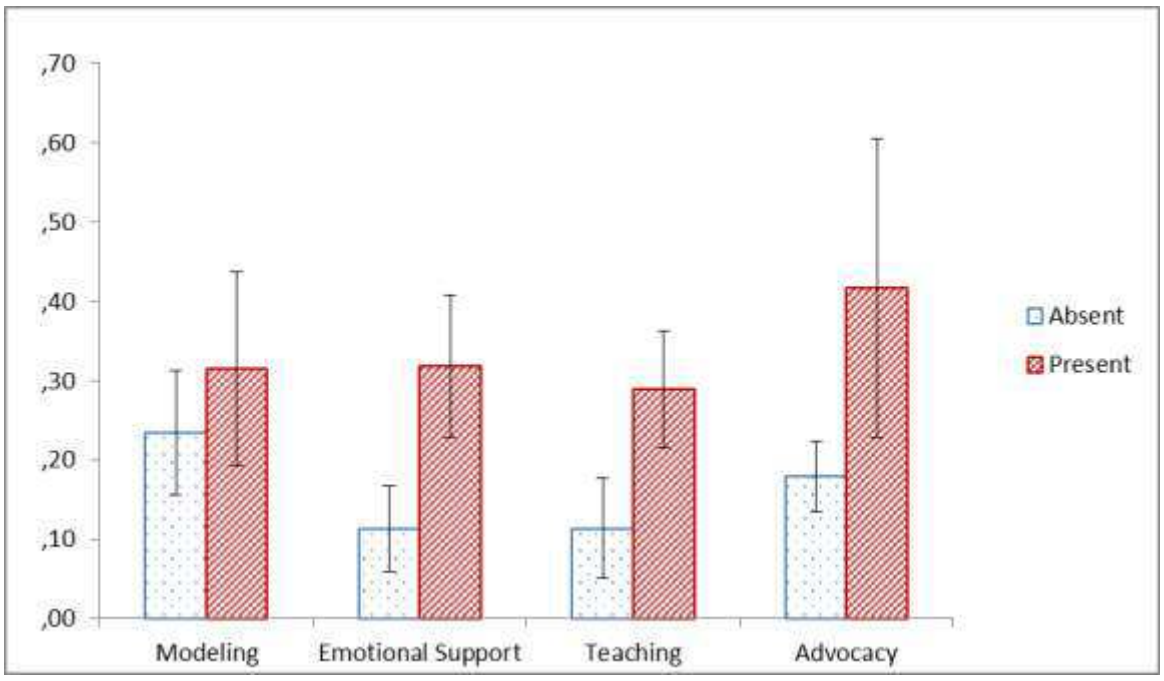


Figure 6



10 Appendices

Appendix 1: Lipsey and Wilson (1998) Codebook
Appendix 2: DuBois et al. (2002) Code Sheet
Appendix 3: Tolan et al. (2004) additional coding

Appendix 1: Juvenile Delinquency Meta-Analysis Coding Manual

REVISED ELIGIBILITY CRITERIA FOR INCLUSION OF A STUDY IN THE DELINQUENCY META-ANALYSIS

1. The study must investigate the effects of an intervention or treatment, broadly defined. In addition to therapeutic type treatments, eligible interventions can include such modalities as incarceration, probation, systems intervention, and the like. Note that the intervention need not explicitly aim to reduce or prevent delinquency. For example, a program to teach delinquents to read would qualify if it met all other criteria even though it was presented as an academic improvement program rather than a delinquency reduction program. The following interventions, however, are specifically excluded: (a) treatments targeted exclusively on substance abuse without attention to any other components of antisocial behavior or outcome variables representing delinquency other than substance use violations; (b) pharmaceutical or medical treatments without significant psychosocial components, e.g., drugs, diet, cosmetic surgery, and the like.

2. The intervention must be applied to a sample that includes juvenile offenders. An offender is defined as a person apprehended by the police, involved with the juvenile or criminal justice system, or identified as having engaged in behavior chargeable under applicable laws, whether or not apprehended or charged. Chargeable offenses include "status" offenses (runaway, truancy, curfew violations, incorrigible, out of parental control) and actions in school and other such contexts that are interpretable as chargeable offenses even if not presented as delinquent behavior, e.g., fighting (assault), damaging school property (vandalism), and the like. A juvenile is defined as anyone under the age of 21 (i.e., age 20.9 or under). If both juveniles and adults are included in the treatment sample, the study is acceptable if the study reports the juvenile results separately or juveniles constitute a majority of the subjects for whom results are reported. Note that if there are any clearly identified juvenile offenders under these definitions in the treatment sample (even one), this eligibility criterion is met.

3. The study must measure at least one quantitative delinquency outcome variable. In addition, it must report results on at least one such a variable in a form that, at

minimum, allows the direction of the effect to be determined (whether the outcome was more favorable for the treatment or control group). If a delinquency outcome is measured but the reported results fall short of this standard, the study will still be acceptable if the required results can be obtained from the author or other sources. A delinquency outcome variable is one that represents, at least in part, the subject's involvement in behavior that constitutes chargeable offenses as defined in 2 above.

4. The study design must involve a comparison that contrasts one or more identifiable focal treatments with one or more control conditions. Control conditions can be "no treatment," "treatment as usual," "placebo treatment," and so forth as long as they do not represent a concerted effort to produce change. Thus, treatment-treatment comparisons are not eligible unless one of the "treatments" is explicitly presented as a form of control condition, e.g., a "straw man" treatment not expected to be effective. When different naturally occurring facilities or groups (e.g., court or probation dispositions) are compared, the study will be eligible only if the different groups are presented as a contrast between a program or intervention of special interest and a control (e.g., "treatment as usual"). For example, a comparison of the pre and post arrest rates for juveniles in each of several probation camps would not be eligible unless it was explicitly presented as a contrast between camps with distinctive programming, e.g., "milieu therapy," and others that followed relatively indistinctive routine and customary practices.

Random assignment designs that meet the above conditions are always eligible under this criterion. One-group pretest-posttest studies are never eligible (studies in which the effects of treatment are examined by comparing measures taken before treatment with measures taken after treatment on a single subject sample). Non-equivalent comparison group designs may be eligible (studies in which treatment and control groups are compared even though the research subjects were not randomly assigned to those groups). To be eligible, however, such comparisons must have either (a) matching of the treatment and control groups prior to treatment on at least one recognized risk variable for delinquency such as prior delinquency history, sex, age, ethnicity, or socioeconomic status; (b) a pre-intervention measure (pretest) for at least one delinquency outcome variable on which the treatment and control groups can be compared; or (c) a pre-intervention measure on at least one recognized risk variable for delinquency (as above) on which the treatment and control groups can be compared. Note that the pre-intervention measures need not show that the treatment and control groups are actually similar, only be capable of showing their degree of similarity (or dissimilarity).

5. The study must be set in the U.S. or a predominately English-speaking country and use juveniles resident to that country. Note that the juveniles need not be English-speaking or "Anglo." A study conducted in the U.S. or Canada with resident Hispanic juveniles, for example, would qualify. In addition, the study must be

reported in English; studies reported in another language will be excluded irrespective of where they were conducted or the nationality of the juveniles.

6. The date of publication or reporting of the study must be 1950 or later even though the research itself might have been conducted prior to 1950. If, however, there is evidence in the report that the intervention under study was applied to the research sample prior to 1945 (i.e., more than five years before the 1950 cutoff date), then the study should be excluded.

ELIGIBILITY CHECKLIST

No	Yes	
____	____	Involves a "treatment," broadly defined, that can be viewed as potentially having some practical benefit for juvenile or society; not restricted to a treatment of solely theoretical interest.
____	____	Involves a comparison that contrasts one or more identifiable focal treatments with one or more control conditions.
____	____	Subjects assigned randomly, matched, or pre-treatment group equivalence available?
____	____	Quantitative outcome data or direction of effect available on at least one delinquency outcome measure.
____	____	Involves juvenile delinquents or subjects committing acts which constitute chargeable offenses.
____	____	Subjects are under the age of 21.
____	____	Study is set in an English-speaking country and reported in English.
____	____	Date of publication is 1950 or later.

STUDY HEADER AND EXPERIMENTAL COMPARISONS

Definition of a study

The “unit” to be coded consists of a study, i.e., one research investigation of defined subject samples compared to each other and the treatments, measures, and statistical analyses applied to them. Sometimes there are several different reports of a single study. In such cases, the coding should be done from the set of relevant reports, using whichever is best for each item to be coded; be sure you have the full set of relevant reports before beginning to code. Sometimes a single report describes more than one study, e.g., a series of similar studies done at different sites. In these cases, each study should be coded separately as if each had been described in a separate report.

Study and Coder Identification

[Note: Variable names for SPSS in brackets, e.g., [ID]; these are not shown in FileMaker and can be ignored for coding purposes.]
_____ Identification number of primary report as assigned in the master bibliography [ID].
____ / ____ /____ Date coded [CodeDate]
_____ Coder's initials (3 letters) [Coder]

CONTEXT SCREEN

Type of publication [SH2] (if multiple, code highest in list; e.g., if dissertation and journal article, code study as journal article).

1 book
2 journal article/book chapter
3 thesis/dissertation
4 technical report
5 conference paper
6 other: _____

Year of publication [SH3] (two digits; estimate if necessary). If you have multiple reports enter the year that corresponds to the report you selected under 'type of publication' above. If there are multiple reports of the same type, use the earliest date. [Eligibility issue- not before 1950]

Senior author's discipline [SH5] (check best one): Note that this question asks about the *senior* author – thus, if more than one author, use discipline of first author.

01 psychology
02 sociology
03 education
04 criminal justice; criminology
05 social work
06 psychiatry; medicine
07 political science
08 anthropology
09 other:
10 cannot tell

Country in which study conducted [SH6]
[Eligibility issue- should be English speaking culture]

1 USA
2 Canada
3 Britain
4 other Commonwealth/English speaking
5 other
6 cannot tell

Role of evaluator/author in the program [SH9] (if more than one, check the highest on the list): [Note: This item is focusing on the role of the research team working on the evaluation regardless of whether they are all listed as authors.]

1 Evaluator delivered therapy/treatment
2 Evaluator involved in planning, controlling, or supervising delivery treatment or Evaluator is designer of program
3 Evaluator influential in service setting but no direct role in delivering, controlling, or supervision
4 Evaluator independent of service setting and treatment; research role only
5 cannot tell

Program age at time of research [SH10] (check best judgment): [Note: If several treatments of different sorts, answer in terms of the treatment to be used in the aggregate experimental comparison, next section. If organization predates treatment, respond in terms of how new treatment is if can assess; if not, indicate how new organization is if can assess. This item is attempting to distinguish between inexperienced, formative, immature programs and those that have been refined and are more mature.]

1 relatively new, e.g., less than two years old or first of relatively few client cohorts
2 established program, in place two years or more, or many client cohorts
3 defunct program, evaluated post hoc
4 cannot tell

Program sponsorship [SH11] (check best one): [Note: Who administers and "owns" the program irrespective of where housed. This is a question of who makes decisions like staffing, changing the program, etc. The first two categories are basically for research and demonstration programs organized by researchers primarily for research purposes. Usually the last three categories are the appropriate choices if the work is done in a service agency even if for research purposes.]

1 demonstration program/treatment administered by researchers for one treatment cohort only
2 demonstration program/treatment run by researchers for multiple treatment cohorts
3 independent "private" program with own facility, staff, etc. (e.g., YMCA, private agency, university clinic)
4 public program, non criminal justice sponsorship (e.g., school sponsored, community mental health, department of social services)
5 public program, criminal justice sponsorship (e.g., police, probation, courts)
6 cannot tell

GROUPS SCREEN

Experimental Comparisons Worksheet

Step 1: Identify all group comparisons in the study. A comparison consists of a configuration in which group differences are or could be tested with t-tests, F-tests, Chi-squares, etc. applied to various dependent measures. Your concern now is with the group comparisons, not the number or nature of dependent measures on which they may be compared (that comes later). For example, one treatment group compared with one control group on six dependent measures is one experimental comparison. The full range of interesting variation on experimental comparisons expected in studies includes the following three possibilities:

(a) Aggregate treatment and control groups. The largest subject groupings on which contrasts between experimental conditions can be made. Often there is only one aggregate treatment group and one aggregate control group, but it is possible to have a design with numerous treatment variations (e.g., different levels) and control variations (e.g., placebos) all compared (e.g., in ANOVA format). These are the groups you will identify on the GROUPS screen.

Step 2: Write in the name/description of <u>each</u> aggregate treatment group and <u>each</u> aggregate control group in the appropriate boxes and, underneath, the number (count) of such groups.
[SH24]: Total number of treatment groups from this study.
[SH25]: Total number of control groups from this study.

Step 3: You will code only one aggregate treatment vs. control comparison plus selected breakouts and post-treatment follow-ups. If there is more than one aggregate treatment group and/or more than one aggregate control group, a selection of which pairing to code must be made as follows:

(a) More than one aggregate treatment group. First, determine if the various treatments are sufficiently similar to combine. This requires that treatment be virtually the same, at least by generic label, for each group, e.g., groups with the same treatment but implemented at different sites or stratified into subgroups that can be recombined into a sensible whole. In such cases, combine the treatment groups into a composite whole if appropriate statistics are available (note: an Excel calculator called “group combo” is available to do the required computations for this in some cases). If statistics for combination are unavailable, select one treatment group to code, as indicated below, and drop the others. Note that if each treatment group has its own distinct control group, separate studies are constituted requiring that each treatment-control pair be coded as independent studies.
If the treatments are distinct, e.g., deliberate experimental variations, and cannot be combined into a relatively uniform composite, then one must be selected as follows:

- If one treatment is clearly the focal concern of the study, with others serving as examples of more conventional approaches, etc., then select the focal treatment.
- If the treatments are parametric variations, e.g., counseling with and without advocacy, then select the most complete or extensive treatment, e.g., the counseling with advocacy. Extensive refers to breadth of services not number of hours of service. This is a subset/superset issue. If one treatment is a subset of another, in the sense of having some but not all of the treatment elements of the other take the superset as the treatment group of interest.
- If the treatments are different, of equal interest to the study, and of equal completeness, then select the one with the largest N. If equal N, select the one that is least unusual and if equal in that regard, make a random choice (coin toss).

(b) More than one aggregate control group, e.g., attention placebo, no control, etc. Select the best control group available to code from the rank order listing below (best listed first):

1) "no treatment" control (control gets no treatment, left alone)
2) placebo control (controls get some attention or sham treatment)
3) treatment as usual control (controls get "usual," handling instead of special treatment, e.g., regular probation or school)
4) "straw man" alternate treatment control not expected to be effective but used as contrast for treatment group of primary interest

If there are multiple groups in any of these categories, combine them if possible and sensible; otherwise, choose the one aimed at the group most similar to the group receiving the treatment of interest. If you still can't choose on this basis, randomly select one group as the control.

If there are no control groups in these categories, i.e., an uncontrolled study or one comparing alternate treatments to each other but not to a control, the study is ineligible for coding. *Be careful, however, not to confuse "treatment as usual" controls, which are eligible, with "treatment-treatment" comparisons, which are not eligible*. If a treatment is a deliberately designed as an "add on" to the conditions the juveniles otherwise experience, then it cannot be considered a control. Treatment as usual is the normal or usual condition of the juveniles at issue. For example, in a study of treatment of probationers, the "usual" treatment is normal probation. Comparison of juveniles on normal probation with those receiving special intensive supervision, extra counseling, or the like would be an eligible study. Also, do not confuse a placebo treatment, which is eligible, with an "alternate treatment" comparison. A placebo treatment is deliberately set up for the purpose of making a particular contrast with treatment, i.e., it has certain characteristics of treatment but lacks the presumed critical ingredient. Alternate treatments, by contrast, are legitimate treatments in their own right, not defined in terms of their role as a

contrast for the focal treatment of interest. Sometimes an alternate treatment is used for comparison with no expectation that it will be effective, i.e., it is a "straw man" treatment perceived ineffective and included for contrast with an identifiable focal treatment of primary interest. In such cases, the alternate treatment control would be eligible-it is virtually a placebo condition.

Reminder: If there are multiple treatments, each paired with its own control group(s), these are coded as separate studies. The above applies only to cases where multiple treatments and/or multiple controls are compared altogether in a single multi-group study.

Step 4: Finally, write the names/descriptions of the aggregate treatment and aggregate control group chosen in the designated places at the bottom of the GROUPS screen. Note: At this point, the one aggregate experimental comparison to be coded has been identified (i.e., one aggregate treatment group compared with one aggregate control group). Only that one aggregate comparison should be considered in completing the remainder of the coding.

GROUP EQUIVALENCE SCREEN

The unit on which assignment to groups was based [SH26] (check best one):

1 individual juvenile, i.e., some juveniles assigned to treatment, some to comparison group (this is the most common case)
2 classroom, facility, etc., i.e., whole classrooms, etc. assigned to treatment, comparison groups
3 program area, regions, etc., i.e., region assigned as an intact unit
4 cannot tell

How subjects assigned to treatment and control groups [SH27] (check best one):
Random or quasi-random:

1 randomly after matching, yoking, stratification, blocking, etc. (This means matched or blocked first then randomly assigned within each pair or block. This does not refer to blocking after treatment for the data analysis.)
2 randomly without matching, etc. (includes also cases such as when every other person goes to the control group)
3 regression discontinuity; quantitative cutting point defines groups on some continuum (this is rare)
4 wait list control or other such quasi-random procedures presumed to produce comparable groups (no obvious differences). [This applies to groups which have individuals apparently randomly assigned by some naturally occurring process, e.g. first person to walk in the door.]

Nonrandom, but matched (control group selected to match treatment group):

5 matched on pretest measures of some or all variables used later as outcome measures (individual level)
6 matched on demographics: big sociological variables like age, sex, ethnicity, SES, (individual level) [Note: If matched on both personal characteristics and demographics call it the former not the latter]
7 matched on personal characteristics, delinquency history, introversion-level, self-esteem, etc. other than dependent variables used later as outcome measures (individual level)
8 equated groupwise; e.g., picking intact classroom of similar characteristics to treatment classroom e.g. mean age of groups are equal.

Nonrandom, no matching (descriptive data regarding the nature of the group differences before treatment must be available for study with this design to be eligible; if initially nonequivalent groups, posttest only, with no information about group similarity, then study is not eligible for coding):

9 originally random or quasi-random but with refusals, exclusions, selections, or other degradations after assignment and before treatment starts amounting to 10 to 15 percent of group or more. [Note: This does not refer to attrition after treatment begins, only between point of assignment and onset of treatment, e.g. groups selected randomly from school roster but many refuse to participate in offered treatment. Treatment drop-out issues are coded elsewhere.]
10 individual selection on basis of need, volunteering, convenience, or some other such factor
11 convenience comparison groupwise, i.e., other available group such as a classroom taken w/o matching or equating (like individual selection but done groupwise)
12 other: _____
13 cannot tell

Confidence of judgment on how subjects were assigned [SH28]:

Very Low	Low	Moderate	High	Very High
1	2	3	4	5
(Guess)	(Informed Guess)	(Weak Inference)	(Strong Inference)	(Explicitly Stated)

Identify all the variables for which comparisons were made between the treatment and control group prior to application of the treatment. These are comparisons that would indicate how similar the treatment and control groups were on some variable(s) after assignment to the respective groups but before treatment was given to the treatment group. Divide these comparisons into two categories:

a) statistical comparisons– variables on which the groups are compared in terms of statistics such as means or proportions, or for which the results of statistical significance testing is reported;
b) descriptive comparisons– variables for which it is reported that there is or is not a difference but no statistics are provided nor any indication of the results of statistical significance testing.

Number of variables statistically compared prior to intervention [SH30]:
Number of variables descriptively compared prior to intervention [SH31]:

General Results of Equivalence Comparisons. [SH29] Select ONE (if both, use statistical).
[Note: For the ratings below, an "important" difference means a difference on most of the variables, or on a major variable, or large differences; major variables are those likely to be related to delinquency, e.g., history of delinquency or other antisocial behavior (chargeable offenses), delinquency risk or prediction, sex, age, ethnicity, SES, family circumstances, temperament.]

Note also that this item is best answered after you make your group equivalence effect sizes (described below) so that you can incorporate the magnitude of the effect sizes into your decision about their importance.
1 no comparisons made

Results of statistical comparison(s):
2 no apparent differences
3 differences exist, but judged unimportant by coder
4 differences exist, judged of uncertain importance by coder
5 differences exist, and judged important by coder

Results of descriptive comparison(s) [if no statistical comparisons made]:
6 negligible differences, judged unimportant by coder
7 some differences judged of uncertain importance by coder
8 some differences, judged important by coder

STATISTICAL COMPARISON WORKSHEET
For each variable identified below on which the treatment and control group were compared prior to treatment (other than pretests on outcome variables) OR on which you can tell equivalence (e.g., if matched on age, etc.) AND for which sufficient data exists, determine the direction of difference and if possible, calculate an effect size. NOTE: you only have to make one effect size for each comparison type (e.g., if you have two measures of age, like average age in years and average grade, you need only make one group equivalence effect size.)

In the case of all male samples, there is no need to make a group equivalence effect size for sex, although you would use this information is judging group similarity and within group heterogeneity below.

Do not include here any comparisons on pretest variables, that is, measures of an outcome (dependent) variable taken prior to treatment (e.g., prior number of arrests in six-month period when number of arrests in six months subsequent to treatment is used as an outcome measure). In such cases the pretreatment ES is coded later as pretest information, not here as group equivalence information. Prior delinquency is a pretest for a delinquency outcome measure if it is in the same form as the posttest (e.g. both court records or both self report but not one of each), measures the same thing, and covers the same time interval (e.g., whether arrested in six-month period). If the prior delinquency IS a pretest, DO NOT code it here. One rule is that it is a pretest if you could compare this with the posttest and get something meaningful.

(a) A variable is only a pretest if it is operationalized exactly like the posttest in all regards except time of measurement. Note especially that for delinquency measures the time period covered must be identical for a pre and post measure to qualify; total prior arrests before treatment is not a pretest for arrests over the six months after treatment.
(b) See codebook for instructions on calculating effect sizes. Be sure the sign of the ES is correct- positive ES favors treatment group, negative ES favors control group.
(c) If there is more than one eligible variable in any of these categories, report on the one that has the most complete information or, in the case of prior delinquency history and typology, the one most relevant to overall delinquency risk.
(d) The variables considered here are the same ones that are eligible for coding in the section on breakouts and should be coded there if available.

Type of Comparison [SC4]

1 Sex
2 Age
3 Ethnicity
4 Prior Delinquency History
5 Delinquency Typology or Risk Level (e.g., type of offender, propensity to commit crime, etc.)

If you have two measures of prior history (like severity and type of offense) use severity as prior history and type as typology if you have no other typology information. If you have all three either throw out type or aggregate it with severity, by averaging the ES values.

Direction Favors [SC5] (Direction of the raw difference on the statistics or description provided):

1 favors treatment group (Tx has fewer males, is younger, has fewer minorities, less delinq history, or less delinquency risk)
2 favors control group (see above)
3 favors neither (exactly the same, reported as no difference, matched)
4 ?? cannot tell

Groups matched on this variable? [SC6] Yes or No

|__|__|__| treatment group sample size for ES calculation [SC1]

|__|__|__| control group sample size for ES calculation [SC2]

|__|__|.|__|__| effect size (two decimals with an algebraic sign in front: plus if favors treatment, minus if favors control) [SC21]
Once you've coded the group equivalence effect sizes, return to the Header file and complete the group equivalence coding.

Similarity rating [SH52]:

Using all the available information, including method of assignment to groups (whether random, matched, etc.), rate the overall similarity of the treatment group and the comparison group, prior to treatment, on factors likely to have to do with delinquency and responsiveness to treatment (ignore differences on any irrelevant factors).

[Note: Greatest equivalence from "clean randomization" with prior blocking on relevant characteristics and no subsequent degradation; least equivalence with some differential selection of one "type" of individual vs. another on some variable likely to be relevant to delinquency, e.g., police referrals for treatment compared with "normal" high school sample.]

[Guidelines: The bottom 3 points are for good randomizations and matchings, e.g., 1=clean random, 2=nice matched. The top three points are for selection with no matching or randomization. Within this bracket, the question is whether the selection bias is pertinent. Were subjects selected explicitly or implicitly on a variable that makes a big difference in delinquency? The middle three points are for sloppy matching designs, degradations, bad wait list designs, and the like. If the data indicate equivalence but the assignment procedure was not random give it a 4 or thereabouts since not all possible variables were measured for equivalence between groups.]

Very similar, equivalent						Very different not equivalent
1	2	3	4	5	6	7

Confidence rating [SH53]:

Very Low	Low	Moderate	High	Very High
1	2	3	4	5
(Guess)	(Informed Guess)	(Weak Inference)	(Strong Inference)	(Explicitly Stated)

____ NA for cannot tell

SUBJECTS SCREEN

CHARACTERISTICS OF SUBJECTS IN TREATMENT GROUP

[Note: LE=law enforcement; JJ=juvenile justice;]

Note: the offense that results in the juvenile entering treatment "counts" as an offense for purposes of this question and the following questions about the juveniles' prior histories.

Predominant level of reoffense risk of treated subjects [SH81] at onset of treatment (check best one):

1 nondelinquents, normal (no evidence of LE or JJ contact or illegal behavior; no identified symptoms or risk factors; regular kids)

2 nondelinquents, symptomatic (no evidence of LE or JJ contact or illegal behavior, but risk factors such as poverty, family problems, school behavior problems, Glueck scale scores, teacher referrals, etc.)

3 predelinquents, minor police contact (no formal probation or court contact or minor self-reported delinquency minor drug infractions, traffic and status offenses, counseled and released, etc.)
4 delinquents (formal probation and/or court adjudication but noncustodial or significant self-reported delinquency, e.g., burglary, property crimes, auto theft; any juvenile who went to court
5 institutionalized, non JJ setting (e.g., mental health in-patient; not just detained pending hearing)
6 institutionalized, JJ setting (e.g., in group home, camp, reform/training school, etc. under court order)

These first six constitute our risk scale; the remaining items are for mixed groups in which no single "type" predominates.

7 mixed, mostly low end of range (nondelinquent & predelinquent)
8 mixed, mostly moderate to high end of range (predelinquent & delinquent/sometimes institutionalized) [Note: This is appropriate if there are offenses for all of the kids.]
9 mixed, full range (nondelinquent through delinquent/institutionalized)
10 cannot tell

Confidence in judgment of level of delinquency (or crime) risk [SH82]:

Very Low	Low	Moderate	High	Very High
1	2	3	4	5
(Guess)	(Informed Guess)	(Weak Inference)	(Strong Inference)	(Explicitly Stated)

____ NA for cannot tell

Number of treated subjects w/ officially recorded priors [SH83]:
Approximately how many of the treatment juveniles have prior offense records (check best one):
1 none
2 some (<50%)
3 most (= or >50%)
4 all (>95%)
5 some, but cannot estimate proportion
6 cannot tell

Predominant type of prior offense reported for treatment subjects [SH84] (check best one):
1 no priors
2 mixed or undifferentiated offenses (you know there are offenses but you do not know what types or the percentage of subjects with each)

3 person crimes (assault, sexual)
4 property crimes (burglary, theft, vandalism)
5 drug/alcohol (possession, sale, public intoxication)
6 status offenses (runaway, truancy, incorrigible)
7 other specific:
8 cannot tell

Number of treated subjects w/ aggressive histories [SH85]: Does the history of the treated juveniles include any suggestion of aggression, violence, assaultive behavior against persons, etc. whether officially recorded or not (check best one):
1 no
2 yes, some juveniles (<50%)
3 yes, most juveniles (= or >50%)
4 yes, all juveniles (>95%)
5 some, but cannot estimate proportion
6 cannot tell

Sex of treated subjects [SH86] or best guess (check best one):
1 no males (>95% female)
2 some males (<50%)
3 mostly males (= or >50%)
4 all males (>95%)
5 some males, but cannot estimate proportion
6 cannot tell

Approx. mean age of treated subjects at time of treatment [SH87](one decimal; 99.9 if cannot tell) [Note: Code best information available even if must estimate, e.g., from grade levels]

How reported? [SH88]How reported/determined (check one used): [Note: Listed in order of preference; if have choice, take higher form in list]
1 median
2 mean
3 mode
4 midpoint of range
5 inference from school grade or other such information
6 not applicable

Predominant ethnicity of treatment subjects: [SH89] more than 60% of juveniles (check best one or best guess):
1 Anglo
2 Black
3 Hispanic
4 other minority

5 mixed (several, but none more than 60%)
6 mixed, but cannot estimate proportions
7 cannot tell

Using above information, how heterogeneous is the treatment group?
[SH90] Overall heterogeneity rating: Based on all the evidence available, how diverse or heterogeneous is the treatment group with regard to delinquency history, demographics, personal characteristics, and conditions relevant to delinquency, etc.? [Note: The issue is one of within group heterogeneity. A highly selective group would rate 1 or 2 and a program that takes all comers would rate a 6 or 7.]

Very Homogeneous (Juveniles quite similar to each other)	1	2	3	4	5	6	7	Very Heterogeneous (Juveniles quite different from each other)

___ cannot tell

Confidence in homogeneity rating: [SH91]

Very Low	Low	Moderate	High	Very High
1	2	3	4	5
(Guess)	(Informed Guess)	(Weak Inference)	(Strong Inference)	(Explicitly Stated)

___ NA for cannot tell

CONTROL SCREEN

WHAT'S DONE TO CONTROL GROUP [SH54]

What the control group receives (select best one): [Note: The difference between 'receives nothing' and 'treatment as usual' hinges on whether or not the two groups have an institutional framework or experience in common, e.g., probation supervision, institutionalization, school.]

1 receives nothing; no evidence of any treatment or attention; may still be in school or on probation etc., but that is incidental to the treatment strategy or client population as defined
2 wait list, delayed treatment control, etc.; contact limited to application, screening, pretest, posttest, etc.
3 minimal contact; instructions, intake interview, etc. ; but not wait listed
4 parole—treatment as usual
5 school—treatment as usual (if treatment delivered in a school setting)

6 probation—treatment as usual(if treatment delivered in a juvenile justice setting)
7 institutionalization—treatment as usual
8 other—treatment as usual
9 attention placebo, e.g., control receives discussion, attention, or dilute version of treatment
10 treatment element placebo; control receives target treatment except for defined element presumed to be the crucial ingredient
11 alternate treatment; control is not really a "control," but another treatment (other than "usual" treatment) being compared with the focal treatment [Such comparisons are not eligible for coding unless the alternate treatment is designed as a contrast to a focal treatment, e.g., a very dilute dose or a "straw man" not expected to perform well.]
12 cannot tell

Overall confidence of judgment on what control group receives: [SH55]

Very Low	Low	Moderate	High	Very High
1	2	3	4	5
(Guess)	(Informed Guess)	(Weak Inference)	(Strong Inference)	(Explicitly Stated)

____ NA for cannot tell

Text box for notes about control group

Describe the character of the control group briefly with particular attention to any experiences they have in common with the treatment group (e.g., "also on probation") and what part of their experience is distinctly different from that of the treatment group (e.g., "in regular institution rather than cottages and doesn't participate in the guided group program").

TREATMENT SCREEN

WHAT'S DONE TO TREATMENT GROUP

Source of clients for treatment [SH56] (check best one): [Note: The issue here is who took the initiative in identifying or choosing subjects for the treatment, e.g., were they identified by teachers or by researchers using the teachers' records?]

1 sought treatment voluntarily ("self-referral," "walk-in")
2 referred/identified by parents, friends
3 referred/identified by non CJ community agency (schools, teachers, mental health, etc.)

4 referred/identified by CJ agency, but "voluntary" (e.g., via police, probation, court, etc.)
5 referred/identified by CJ agency, but participation mandated (e.g., by court, terms of probation, institution). [Assume it is mandatory if it is a CJ agency unless there is specific information that it is voluntary. Don't override a specific statement that it's voluntary even if you presume, there is some coercion.]
6 referred/identified by multiple sources, none predominates
7 solicited or arranged by researcher
8 other _____
9 cannot tell

Type of treatment: Link to Service Codes Screen

Overall confidence in judgment about type of treatment: [SH59]

Very Low	Low	Moderate	High	Very High
1	2	3	4	5
(Guess)	(Informed Guess)	(Weak Inference)	(Strong Inference)	(Explicitly Stated)

____ NA for cannot tell

Who administers treatment [SH61] (check best one):

1 criminal justice or juvenile justice personnel (e.g., police, probation officer, judge, etc.)
2 school personnel (e.g., teachers, principals)
3 mental health personnel (public agency)
4 mental health personnel (private agency, counselors, etc.)
5 non mental health professionals, counselors, consultants, etc., e.g., vocational counselors
6 laypersons, e.g., volunteers, college students, ex-delinquents
7 researcher/research team
8 other: _____
10 mixed, multiple personnel (contact with more than one treatment delivery person & none is clearly focal). Do not use this option when different subjects are seeing different types of personnel. In those cases, select a focal personnel type.
11 cannot tell

Format of treatment sessions [SH62](check best one; if mixed, check predominant category):
(Note: The primary emphasis of this question is on who was present with the juvenile during treatment, emphasis on number of providers present is secondary)

1 juvenile alone (self-administered treatment, e.g., bibliotherapy) [This refers to a treatment in which nobody else is present. If it is restitution performed in a group it does not belong here but if a juvenile is sent out to do something (like get a job) it goes here.]
2 juvenile and provider, one on one
3 juvenile group, one or more providers
4 juvenile with family/parents, one or more providers
5 parents only, juvenile not present
6 teachers, probation officers etc. only; juvenile not present
7 mixed; no single format predominates
8 other: _____
9 cannot tell

Nature of treatment site: [SH63] site on which treatment generally delivered (check best one in each set): [Note: Customary treatment location irrespective of who administers treatment. If restitution is the treatment, the site will be mixed, none predominates.]

1 Public facility (i.e., owned and operated by city, county, state, federal government body), JUSTICE-ORIENTED, e.g., probation dept, police station, reform school
2 Public facility (i.e., owned and operated by city, county, state, federal government body), NOT JUSTICE-ORIENTED, e.g., school, dept. mental health
3 Private facility, e.g., YMCA, private counseling agency, university (even if state university)
4 mixed, none predominates
5 other: _____
6 cannot tell

Custodial/residential facility? [SH64] e.g., camp, reformatory, Psychiatric hospital, halfway house, foster home, etc.

1 yes
2 no
3 mixed, neither predominates
4 cannot tell

Formal setting? [SH65] (e.g., office, classroom, institution, laboratory, etc.)

1 yes
2 no, informal, e.g., outdoors, streets, juvenile's home, etc.
3 mixed, neither predominates
4 other: _____
5 cannot tell

SERVICE CODES SCREEN

Treatment description [SH100txt]

Relationship of Juveniles in Treatment to the Juvenile Justice System [SH100]

The purpose of this item is to capture the status of the juvenile at the time treatment was actually received. Juvenile justice supervision means that they are officially supervised while on probation, in a residential/custodial facility, or on parole/aftercare and can be sanctioned by the JJ authorities if they fail to comply with the terms of that supervision. A juvenile is not under the authority of the JJ system if they are not being monitored on an on-going basis by JJ authorities. Non-JJ supervision can include juveniles that were routed to services via the JJ system (diversion), but are participating in the services without official JJ supervision.

Yes, juveniles under JJ supervision (under the authority of the JJ system)when they received the treatment

On probation (under probation supervision but not in custodial institution nor aftercare/parole after a term in a custodial institution).

1 on probation, in community (or no indication that not). Describe:
2 on probation but in a residential or partially residential setting, e.g., day treatment, probation camp. Describe:

In a juvenile justice custodial institution, e.g., training/reform school, borstal, detention center, juvenile correctional institution.

3 ”regular” juvenile correctional institution (or no indication that not). Describe:
4 alternative or special form of custodial institution, e.g., cottage format, psychiatric correctional ward. Describe:

On JJ supervised parole of aftercare after a term in a custodial institution (after incarceration).

5 nonresidential JJ parole or post-custodial aftercare. Describe:
6 partial residential JJ parole or post-custodial aftercare, e.g., day treatment program. Describe:
7 fully residential JJ parole or post-custodial aftercare, e.g., group home, halfway house. Describe:

Any other form of JJ supervision or under JJ authority but cannot tell which of above is applicable.

8 other JJ supervision. Describe:

No, juveniles not under JJ supervision when treatment received (through some route such as diversion by law enforcement or juvenile justice personnel, and are not under JJ supervision while in treatment.)

Note: If juveniles initially involved with police or juvenile justice system but then diverted away from official JJ processing and released or sent to a community program, note this in the write-in space for description for the option to which it applies. Such a situation may involve the threat of JJ processing if treatment is not completed but the juvenile will not actually be under JJ supervision at the time of treatment following the diversion.

9 in the community with no apparent constraints or residential program arrangement. Describe:

10 in a non-JJ partially residential setting, e.g., non JJ day treatment program, alternative school. Describe:

11 in a non-JJ fully residential setting, e.g., group home, foster care. Describe:

12 other non JJ situation. Describe:

All other or cannot tell which of the above apply.

13 Cannot tell. Describe:

Treatment Components

Identify all the treatment components, elements, activities, experiences, etc. reported as part of the intervention. Note that to qualify, a component should be something the treatment group receives that the control group does not receive. Use the following rating scale for each reported component. At least one component must be rated for every intervention but as many components can be rated as needed to describe every distinct element reported.

Some items are listed multiple times and are indicated with a similar superscript. Although an item may be listed under several categories, it should only be rated one time for each intervention. Items that are in bold type are considered "brand name" interventions. These should only be chosen if mentioned specifically by name within the study report(s). If the treatment description sounds like it has all or most of the components of a particular "brand name" intervention, but it is not specifically called by that name, place it in the "similar to" category.

It is important to assign a code to **all** treatment components mentioned for each intervention using the numerical scheme below. Initially you should assume that each such component will receive a rating of "1," like "1" was a checkmark to check off every item present. *However*, if there is any indication in the study report(s) that

one or more components are of lesser scope or importance than others, then those secondary items should be coded "2." A component might be identified as secondary in this sense because:

a) it is clearly a subcomponent of something else (e.g., role-playing as one of several parts of a attitude change session) or there is a broad program type to be coded "1" (e.g., interpersonal skills building) and the component is only one aspect of that (e.g., anger management exercises);
b) it is provided to only a subset of juveniles or only occasionally in contrast to other components provided to all juveniles or on all occasions (e.g., a service that some juveniles are referred to only if they need it while others are provided to all)
c) some other distinction is made that shows that the component is not of equal importance, stature, or scope as others that are coded "1."

If there is no basis for distinguishing any components as having less importance, scope, stature, etc. than any other, code all as "1." If you have some reason to doubt that all the components are at the same level, but a clear determination cannot be made about which should be coded "1" and which "2," then code all the uncertain components as a "9."

1. treatment component with no indication that it is a subcomponent, of less scope, provided to fewer juveniles, etc. than any other component
2. a treatment component that is a subcomponent, of less scope, provided to fewer juveniles, etc. than some other component
   one of a set of components that may be at different levels ("1" vs "2" above) but it is uncertain which is which (i.e. cannot clearly and comfortably determine if a component is a "1" or "2")

JJ or CJ-type Treatment Elements

[tc1]	probation, regular (compared to no probation supervision)
[tc3]	parole/aftercare, regular (compared to no parole/aftercare supervision)
[tc5]	institutionalization, regular (jail, detention center, prison, etc. compared to no institutionalization)
[tc7]	early release from institution, probation/, or parole (shortened sentence)
[tc8]	furloughs from custody (e.g., family visits, field trips without JJ staff members)
[tc123]	work release program (e.g., work in the community while still incarcerated)
[tc124]	work program (work in the institution while still incarcerated)
[tc9]	**intensive supervision** or monitoring, **reduced caseload**, smaller units, more frequent drug screens

[tc10] community monitoring (e.g., sex offender registry, electronic bracelet)
[tc11] drug court (e.g., more lenient sentencing to substance abuse treatment in closed facility)
[tc12] prison visit, not overnight (e.g., scared straight, etc.)
[tc13] short term "shock" incarceration (juvenile stays overnight at least 1 night)
[tc14] deterrence threat (e.g., straight talk with police officers, “lecture and release”)
[tc137] Teen Court, type of alt. sentencing & peer review/sentencing format
[tc15] military style “boot camp” (relatively short term)
[tc16] restitution, fines or payment/service to victim or victim’s family
[tc17] restitution, community service (e.g., landscaping, hospital, nursing homes, etc.)
[tc138] restitution, contact with victim (e.g., apology letters, apology in person)
[tc18] **diversion** specifically stated as a descriptor of the program
[tc2] alternative to probation (would be on probation but something else instead)
[tc4] alternative to institutionalization (would be institutionalized but something else instead)
[tc6] alternative to parole/aftercare (would be on parole/aftercare but something else instead)
[tc122] receives treatment/service program instead of JJ supervision
[tc125] receives probation instead of greater supervision, e.g., institutionalization
[tc136] receives informal probation instead of greater supervision, e.g., regular probation, institutionalization
[tc19] other

Residential Components

[tc20] psychiatric facility
[tc21] **teaching family home**
[tc21s] similar to teaching family home
[tc139] emergency shelter/shelter house
[tc22] group home; foster parents
[tc23] wilderness camp, short term– two weeks or less in camp (e.g. Outward bound)
[tc118] wilderness camp, not short term– more than two weeks
[tc15] boot camp
[tc25] other camp
[tc26] residential drug treatment

[tc27] boarding school / residential training school, (cottage model, small scale/disaggregated)
[tc28] **guided group interaction**, in a residential setting (e.g., offenders determine rules & punishment)
[tc28s] similar to guided group interaction
[tc111] **positive peer culture**, in a residential setting (e.g., members are responsible for themselves as well as others and serve as catalysts for helping others and advancing the group)
[tc111s] similar to positive peer culture
[tc29] **therapeutic community**
[tc29s] similar to therapeutic community
[tc30] **milieu therapy**
[tc30s] similar to milieu therapy
[tc31] other

Educational Components

[tc135] school-based: program provided in regular school setting
[tc32] special classes or educational field trips
[tc33] continuation/additional school, (not employment related)
[tc34] tutoring, or current level of education (not employment related) by whom?
[tc35] remedial education, (not employment related)
[tc120] alternative school, as alternative for regular (e.g., public) school
[tc160] educational testing
[tc140] assigning homework
[tc141] teaching juveniles study techniques
[tc142] academic monitoring (e.g., monitoring homework, academic performance, attendance, etc.)
[tc161] computer classes (academic-separate from vocational)
[tc36] other

Counseling Components

[tc37] individual counseling, therapy, psychotherapy, guidance; by whom?
[tc38] group counseling, therapy, psychotherapy; by whom?
[tc127] group counseling, led by a facilitator but not necessarily "talk therapy" (e.g., facilitated discussions)
[tc112] **guided group interaction**, (nonresidential)
[tc112s] similar to guided group interaction(nonresidential)
[tc113] **positive peer culture** (nonresidential)
[tc113s] similar to positive peer culture (nonresidential)
[tc114] **multi-systemic therapy**
[tc114s] similar to multi-systemic therapy

[tc143] client-centered therapy
[tc40] family counseling, family systems, functional family therapy, etc. (w/whole family or juv and parent)
[tc144] multi-family groups, (e.g., "family group" participates in counseling as a whole along with other families
[tc41] parent counseling without juvenile, individual
[tc42] parent counseling without juvenile, parent groups
[tc43] [3]drug/alcohol counseling (see also Drug and Alcohol Components)
[tc44] casework: support/services provided by caseworker (not case manager) interceding with others, helping juvenile, etc. ("all-purpose")
[tc145] in home counseling, counseling takes place in the home of the juvenile or family
[tc45] mediation (counselor mediates/arbitrates between parties in conflict or victim and offender)
[tc46] [4]recreational therapy, (see also Recreational Components)
[tc47] **reality therapy**
[tc146] sex offender counseling
[tc48] crisis counseling, response (e.g., come out to house to intervene)
[tc119] non-specific counseling (not otherwise identified)
[tc49] other

Recreational Components

[tc46] recreational therapy
[tc121] recreation (non-specific)
[tc51] fitness programs (e.g., weights, sports--not for competition, increased exercise)
[tc52] sports, athletics, or athletic events
[tc53] parties, games, recreational outings, field trips (other than educational)
[tc147] adventure-based activities, ropes course, canoeing, etc.
[tc54] arts & crafts, drama, music, dance activities, games, etc. (groups and individually)
[tc55] other

Interpersonal/Personal Skill Components

[tc56]	interpersonal skills building (e.g., communication skills, role playing, assertion training)
[tc57]	resisting group pressure, responding to persuasion
[tc58]	peer/group interaction (meetings, discussions, activities)
[tc59]	mentor provided for juvenile (peer, volunteer, layperson, “big brother”)
[tc60]	juvenile served as mentor as part of tx
[tc61]	moral education, training; religious or spiritual program
[tc62]	interpersonal problem solving, conflict resolution, decision making
[tc148]	personal/self development training (e.g., self esteem building, focusing on indiv. strengths, self-awareness, leadership, goal setting, etc.)
[tc63]	anger management (other than cognitive behavioral); stress management, (see also cog anger mgmt)
[tc64]	other

Cognitive Skills / Cog Restructuring Components

[tc115]	**cognitive/behavioral** intervention (overall focus on altering irrational thinking and behavior)
[tc115s]	similar to cognitive/behavioral intervention
[tc65]	cognitive restructuring (monitoring automatic thoughts, correcting distortions/thinking errors, etc.)
[tc66]	cognitive anger management (hassle logs, identify triggers, use self-statements and anger reducers, etc.)
[tc67]	moral reasoning; empathy & victim impact (moral dilemmas; perspective taking; empathy for victim)
[tc68]	attitude change, accepting authority & rules, new attitude towards law, court, police, peers, etc.
[tc69]	relapse prevention plan; interventions for lapses; high-risk situation planning
[tc70]	other, describe

Behavioral Components

[tc71]	behavioral contracting, contingency management; behavior modification; (e.g., rewards; shaping of specific behaviors; reinforcement for desired behaviors)
[tc72]	behavior modification (e.g., rewards, shaping, reinforcement of behaviors, etc.)
[tc73]	punishment, discipline (e.g., segregation, privileges taken away, denial of family visits)
[tc74]	token economy – tokens earned, redeemable for privileges, goods, etc.
[tc75]	learning by modeling
[tc76]	desensitization, exposure+response prevention, flooding
[tc77]	relaxation training (e.g., deep breathing, counting backward, imaging of peaceful scenes)
[tc78]	meditation (mindfulness therapy, living in the moment, yoga, transcendental meditation)
[tc149]	role playing (non-specific or a general activity, not a technique used with another component)
[tc79]	anger reducing techniques (e.g., push-ups, time-outs, walking around) –(see also cognitive anger mgmt)
[tc80]	other

Employment Components

[tc81]	remedial education, employment related; any functional education (literacy, GED, arithmetic)
[tc82]	tutoring (one on one), teaching machine, help to achieve academic success (employment related)
[tc116]	continuing education (employment related) such as special or advanced classes
[tc83]	employment; supervised group work program
[tc128]	employment; job placement for individual juveniles
[tc84]	career counseling, (career exploration, job readiness, job searching skills, interview skills)
[tc85]	job training -- learning new job content, trade, specific skills (e.g., welding, construction, computer)
[tc150]	vocational field trip (separate from educational or recreational field trip)
[tc151]	non-paid work service (e.g., community service not in conjunction with restitution, etc.)
[tc162]	computer classes (vocational—separate from academic)
[tc186]	other

Life Skills/Needs Components

[tc87]	personal management (attendance, housing issues, time/money management skills)
tc88]	managing daily life problems (problem solving, social/moral reasoning, balancing responsibilities)
[tc89]	challenge programs, short term (e.g. survival training, outward bound)
[tc90]	parenting / family skills for parent of target juvenile; (parent effectiveness training alone or with juvenile)
[tc152]	provides necessities (e.g., clothes, transportation, food, etc.)
[tc91]	health-related prevention (pregnancy, STD)
[tc153]	health education (e.g., personal hygiene, nutrition, etc.)
[tc154]	legal education (juveniles learn about the judicial system and judicial processes)
[tc92]	other

System-Oriented Components

[tc93] advocacy on behalf of youth (must be clearly identified as all or part of the treatment program)
[tc94] consultation, assistance to schools/agencies responsible for juveniles' welfare
[tc95] special training for service providers, (school staff, counselors, probation officers)
[tc96] facilitative assistance for service providers, other than training (group discussions, information sharing)
[tc97] parents of juvenile offender receive skill building intervention other than parenting skills (w/o juvenile)
[tc155] regular contact with parents (parental involvement)
[tc98] outreach workers, **streetworkers** (service personnel working with gangs, schools, etc.)
[tc99] other

Drug and Alcohol Components

[tc100] drug, alcohol education
[tc43] [3]drug, alcohol counseling/therapy, (AA or NA)
[tc156] drug testing (conducted either on a regular or random basis)
[tc102] other, (see also Behavioral Components)

Pharmacological, Medical, Biological Components

[tc103] psychiatric intervention (e.g., access to psychiatrist for evaluations & prescriptions)
[tc157] medical/emergency service
[tc104] change in behaviors, diet, medication, sleep, etc., describe:
[tc105] physical examination and necessary treatment (medicine)
[tc106] other

Multimodal Components

[tc107]	service brokerage: evaluation/assessment of service need, referral to treatment; provided by an agency
[tc158]	psychological assessment (separate from assessment for service brokerage)
[tc159]	individualized treatment plans provided for juveniles
[tc108]	multimodal service – program tailored to juveniles receiving multiple tx components
[tc109]	case management (case manager identifies needs, oversees services by multiple agencies, etc.
[tc110]	other

All Other

[tc117 & tc129-tc134] any other treatment component, element, technique, etc. identified in study report(s) and not coded above. Describe with at least moderate detail if possible:

IMPLEMENTATION SCREEN

TREATMENT IMPLEMENTATION/STRENGTH/INTEGRITY

[Note: For this item and the next three use "facts" if available, otherwise "format". Make an informed guess about the amount and frequency of contact whenever possible. Even if the guess is inaccurate, it will help give an order of magnitude estimate for the analyses' Assume that a counseling session and a school period are probably each an hour long.]

Approximate duration of treatment in WEEKS [SH68] from first treatment event to last treatment event. Include treatment received by treatment subjects up to the time of posttest measurement. Divide days by 7 and round; multiply months by 4.3 and round. Code 999 if cannot tell. Estimate for this item if necessary and if you can come up with a reasonable order of magnitude number. If no other information is provided in the study, you can assume that probation lasts 6 months and crisis counseling lasts 2 weeks.

Determined by [SH69] (select one):

1 facts (data about how long clients in treatment, e.g., average client attended 7.3 weeks)
2 format (standard package or plan without information on actual participation, e.g., a ten-week program)
3 other estimate (e.g., coder's best guess)

Frequency of treatment event/contact [SH70] (check best one) [Note: This refers only to the element of treatment that is different from what the control group receives. Estimate for this item if necessary and if you can come up with a reasonable order of magnitude number.]

1 continuous (e.g., milieu therapy, residential program, pharmaceutical therapy, parent effectiveness training)
2 daily contact (not 24 hours of contact per day but some treatment during each day, perhaps excluding weekends)
3 2-4 times a week
4 1-2 times a week
5 less than weekly
6 cannot tell

Determined by [SH71](select one): (for continuous treatments assume format unless have specific information about discrepancies from the prescribed format)

1 facts (data)
2 format (standard package/plan) [code continuous treatments here]
3 other estimate (e.g., coder's best guess)

Approximate mean HOURS of contact per WEEK [SH72] (888 if institutional): actual contact time between juvenile and provider or treatment activity per week per juvenile if reported or calculable (Round to one decimal place. Code 888 for institutional residential, or around the clock program; code 999 if not available) [Note: Estimate for this item if necessary and if you can come up with a reasonable order of magnitude number.]

Determined by [SH73](select one):

1 facts (data)
2 format (standard package/plan) [code continuous treatments here]
3 other estimate (e.g., coder's best guess)

Approximate mean HOURS of TOTAL contact [SH74] over full duration of tx: contact between juvenile and provider or treatment activity over full duration of treatment per juvenile if reported or calculable (Round to whole number. Code 8888 for institutional, residential, or around the clock program; code 9999 if not available) [Note: Estimate for this item if necessary and if you can come up with a reasonable order of magnitude number. No decimals here, whole numbers only.]

Determined by [SH75](select one):

1 facts (data)
2 format (standard package/plan)
3 other estimate (e.g., coder's best guess)

Overall confidence in estimates of treatment contact: [SH76]

Very Low	Low	Moderate	High	Very High
1	2	3	4	5
(Guess)	(Informed Guess)	(Weak Inference)	(Strong Inference)	(Explicitly Stated)

___ NA for cannot tell

Evidence of uncontrolled variation in implementation? [SH77]

Based on evidence or author acknowledgment, was there any uncontrolled variation or degradation in implementation or delivery of treatment, e.g., high dropouts, erratic attendance, treatment not delivered as intended, wide differences between settings or individual providers, etc. (check best one): [Note: This question has to do with variation in treatment delivery not research contact. E.g., there is no “dropout”

if all juveniles complete treatment even if some fail to complete the outcome measures; degradation does not mean attrition per se. Implementation and delivery of treatment to the treatment group partly overlaps the research methodology attrition issue but also includes other aspects involving the treatment itself. Assume that there is no problem if one is not specified and the format seems reasonably structured.]

1 yes (describe: _____)
2 possible (describe: _____)
3 no, apparently implemented as intended
4 cannot tell

Taking all evidence into consideration, rate the intensity of the treatment along the two dimensions below:

Rate amount of meaningful contact [SH78] between subject and treatment (frequency, duration). Amount of meaningful contact between juvenile and treatment (frequency, duration): [Note: Use the number of hours of contact to determine whether the treatment falls into the bottom, middle, or high end of the scale and then adjust the rating according to the meaningfulness of the contact. Try to reflect any slippage between format of treatment and actual amount of contact. Fifteen hours of basketball would rate lower than fifteen hours of counseling because there is less contact with the change agent. A total institution experienced for a long time would rate a “7", a two week wilderness program or a 10 week, once a week crisis intervention program would rate about a “4", high slippage and low participation would yield a rating of “I” or “2". A 2 hour per day program would be about a 6 which would be moved down if there is lots of slack time. Fifteen minutes per week would be about a 1; an hour per week or less would be a 2 or 3.

Trivial 1 2 3 4 5 6 7 Substantial

___ cannot tell

Rate intensity of typical tx event [SH79](involving, emotional, etc.)
Intensity of typical treatment event; how involving, emotional, memorable, etc. per contact irrespective of amount of contact: [Note: Intensity relates to the likelihood that this treatment will cause a psychological change or emotional reaction in the juvenile whether therapeutic or not. Scared straight or a wilderness program would rate a “6" or “7", standard counseling would rate somewhere between “3" and “5", and a boy's club after-school basketball program or informal probation would rate somewhere between “1" and “3".]

Weak 1 2 3 4 5 6 7 Strong

___ cannot tell

Overall confidence in treatment ratings: [SH8o]

Very Low	Low	Moderate	High	Very High
1	2	3	4	5
(Guess)	(Informed Guess)	(Weak Inference)	(Strong Inference)	(Explicitly Stated)

____ NA for cannot tell

Dependent Variable Coding Sheet (DV)

For the aggregate experimental comparison coded on this sheet, identify the dependent (outcome) variables on which treatment vs. control group comparison could be made (whether actually made or not) distinguishing delinquency vs. nondelinquency measures. If it is hard to decide whether a measure reflects delinquency or not, err on the side of calling it a nondelinquency measure so that the delinquency measures used in the analyses will be fairly unambiguous. Each dependent variable represents a contrast between two groups often reported as a test of significance.

Exclude variables that reflect only the degree of implementation of the intervention. Exclude variables that do not apply to the entire aggregate comparison, e.g., measures that subdivide categories of another measure such as single vs. multiple offenses only for those that recidivate. Also exclude variables that do not represent the status (behavior, attitudes, etc.) of the juveniles in the treatment and control groups but rather the status of others, e.g., teachers, parents, juveniles outside the experiment. Note that it is okay for teachers, parents, etc. to be the primary treatment recipients (e.g., parent effectiveness training) but dependent variables are nonetheless only coded for the subsequent status of the juveniles involved (e.g., children of those parents). Note also that it is okay for a dependent variable to represent the observations, opinions, etc. of someone other than the juvenile so long as it is something about the juvenile on which they are reporting (e.g., parent opinion about whether the juvenile has improved).

If the same variable is used repeatedly for follow-up, etc. count it only once. Otherwise, list every dependent variable that can be identified as having been used in the study irrespective of how much information is available on it. Write in a brief label for each below:

DELINQUENT BEHAVIOR OUTCOME MEASURES (LIST ALL)

[Definition: Delinquency outcome measures are those that index the degree of criminal or delinquent behavior (constituting at least one chargeable offense). Direct reports of criminal/ delinquent behavior are always included here whether self-

report from the delinquent or records from police, probation, courts, etc. Also included here are other reports of delinquent behavior such as some school or teacher reports, e.g., having to do with disciplinary actions related to (chargeable offenses). The key factor in the delinquency vs. nondelinquency decision are (a) the measure has to do with behavior; non-behavioral constructs, e.g., attitudes, personality trait measures, etc., should be classified as nondelinquency; (b) the activity involved is officially defined delinquency, or related, or else is antisocial behavior in the sense of causing clear harm to persons, property, or self.]
Verbal tags: _____

On Codesheet DM, code each of the above variables for which some treatment group vs. control group comparison can be made, even if only a statement of nonsignificance, no difference, or direction of effects. Code only those DVs for which there is a statement of the direction of the effect even if that statement is that there was no significant difference. Place a checkmark on the list above beside each variable selected for coding. [Note: There will be four types of dependent measures: those that were measured but not mentioned (lost), those that were mentioned with no statement of results, those that were mentioned with a statement of significance or direction, and those that provide enough information to calculate an effect size. All but the first category should be listed here; all in the third and fourth categories should be coded.]

For status offenses (those that are only offenses because the perpetrators are minors, e.g., runaway, truancy, curfew, incorrigible) it is a delinquent behavior if it is presented as an offense in a law enforcement framework (e.g., police or court records), but is a non-delinquent behavior if it is presented in a non-law enforcement framework (e.g., school records). Fighting or other clearly antisocial behaviors (chargeable offenses) (extorting money, beating up fellow students, etc.) are delinquent regardless of the framework in which they are presented. Indicate the appropriate numbers below:

|__|__| Number of delinquency variables selected for coding
|__|__| Number of delinquency variables omitted
[Note: These two values should sum to the total number of variables on the above list. Do not skip this; it is important.]

NONDELINQUENCY OUTCOME MEASURES (LIST ALL):
[Definition: Nondelinquency outcome measures are all those that remain after any delinquency outcome measures are coded on the “delinquency behavior outcome measures codesheet” according to the definitions on that codesheet.]
Verbal tags: _____

On Codesheet NM, code each of the above variables for which some treatment group vs. control group comparison can be made, even if only a statement of

nonsignificance, no difference, or direction of effects. Code only measures representing the behavior, attitudes, perceptions, etc. of juveniles, not measures of the behavior, etc. of others, e.g., teachers, parents, etc. even if they are the recipients of the treatment. Place a checkmark on the list above beside each variable selected for coding. Indicate the appropriate numbers below:

|__|__| Number of nondelinquency variables selected for coding

|__|__| Number of nondelinquency variables omitted

[Note: These two values should sum to the total number of variables on the above list.]

Delinquency Variables

Code a separate screen for each delinquency outcome measure for which the aggregate treatment and control groups can be compared on the first wave of post-treatment outcome. (Subsequent waves and breakouts for this aggregate comparison are coded on separate attachments to be appended to this sheet). Delinquency outcome measures are those that index the degree of criminal or (delinquent) behavior. Direct reports of criminal/delinquent behavior are always included here whether self-report from the delinquent or records from police, probation, courts, etc. Also included here are other reports of delinquent behavior such as some school or teacher resorts, e.g., having to do with disciplinary actions related to delinquent behavior. The key factor in the delinquency vs. nondelinquency decision are 1) the measure has to do with behavior; non-behavioral constructs, e.g. attitudes, personality trait measures, etc., should be classified as nondelinquency; 2) the activity involved is officially defined delinquency, or related, or else is antisocial behavior in the sense of causing clear harm to persons, property, or self.

Type of delinquency/recidivism represented [D1] by this measure (what's counted, irrespective of source of information and authors' label or description of the measure) (check best one):

1 antisocial behavior, not specifically restricted to criminally delinquent acts
2 unofficial delinquent behavior, e.g., from self or observer's report
3 school disciplinary actions relating to delinquent/antisocial behavior
4 arrests or police contacts
5 probation contact, violations, actions, etc.
6 court contact, actions, petitions, convictions, appearances, etc., excluding institutionalization
7 parole contact, violations, action, etc., excluding reinstitutionalization
8 institutional disciplinary actions (relating to delinquent/antisocial activity)
9 institutionalization or reinstitutionalization
10 catchment area crime/arrest rates (Treatment for entire area)
11 catchment area JJ indicators, e.g., probation, court, parole events
12 other: _____
13 cannot tell

Definitional boundaries for measure [D2] (check best one):

01 all “offenses” included (except, perhaps, traffic offenses)

Restricted by type

02 substance abuse only
03 property crime only
04 person crimes only
05 status offenses only
06 criminal offenses only, i.e., all but status offenses
07 other

Restricted by severity

08 only major/felony
09 only minor/misdemeanor
10 other severity restriction
11 other type of restriction: _____
12 cannot tell

Elements reported in measure: [D3] Elements reported in this delinquency measure irrespective of type incident and reporting source (check best one):

1 global dichotomy or polychotomy (e. g., offended or recidivated, yes/no)
2 summed dichotomous (e.g., sum of yes/no on list of specific offenses)
3 frequency or rate, (count of incident; incidents per 1000 persons)
4 severity (seriousness rating or index)
5 event timing (e.g., days without recidivism; time to first offense)
6 proportion or amount of time in custody, under supervision, etc.
7 rating of amount of delinquency, severity, change, etc. (e.g., therapist rating of extent delinquent behavior improved)
8 more than one of above elements combined in composite measure
9 other: _____
10 cannot tell

Source of delinquency data [D4] (check best one):

Self report

1 paper & pencil
2 personal interview
3 telephone interview
4 other: _____
5 cannot tell

Other reports

06 family
07 peers
08 teacher(s)
09 therapist/service provider
10 other: _____
11 cannot tell

Records

12 school
13 police
14 probation
15 court
16 custodial institution
17 regional crime statistics
18 other: _____
19 cannot tell
20 any other: _____
21 cannot tell which of above categories

Properties of this measure demonstrated, reported, or cited (check all that apply):

Properties demonstrated, validity: [DN1]
Properties demonstrated, reliability: [DN2]
Reliability coefficient: [DN2R]magnitude of coefficient, if given (-99 if missing)
Properties demonstrated, sensitivity:
[DN3]sensitivity/responsiveness/discriminant ability [i.e., indication that measure capable of responding to treatment effect]
Properties demonstrated, none: [DN4] none of above

Treatment-test overlap: [DN5] Rate the extent to which the treatment content overlaps or resembles the content of this measure, e.g., as in "teaching the test." At one end of the continuum are measures that are virtual duplicates of the treatment, e.g., a behavioral treatment that reinforces a specific list of behaviors and an outcome measure that counts how often those same behaviors are performed. At the other end of the continuum are measures that have virtually no content similarity to the treatment, e.g., a treatment of insight-oriented counseling about family relations and an outcome measure of math grades in school. This is not a question about the extent to which the treatment caused the dependent variable. The question concerns the content of the treatment not the plausibility of the hypothesized causal relationship. The topic area of the treatment in relation to the topic area of the measure determines the general category. Use the 1-3 range for treatments and measures of generally different content and involving different activities; use 3-5 for those situations like general counseling and delinquency measures where discussion of delinquency may well have been part of the treatment content, giving topic overlap, but the activities of treatment (talking about delinquency) are different from those in the measure (committing delinquency). Use the 5-7 range for fairly clear overlap in both topic area and activity, e.g. substance abuse treatment involving role playing resistance to peer pressure and actual substance abuse incidents as an outcome measure. Within these ranges, adjust for the degree of overlap according to the specifies of the individual case.

Rate this measure for treatment-test content overlap:

Very Low Overlap 1 2 3 4 5 6 7 Very High Overlap

Social desirability bias: [DN6] Rate the extent to which this measure seems susceptible to a social desirability response bias, that is, the extent to which the respondents are (a) able to recognize what response "looks good," (b) may be motivated to "look good," and (a) are able to exaggerate the response in the direction of "looking good." Note that you are not to rate how much social desirability bias you

think actually occurred, only how susceptible you think the measure might be. At one end of the continuum would be measures based on objective procedures administered by impartial others, e.g., random surprise urinalysis for drug testing. At the other end of the continuum would be the juvenile's own reports made to someone with authority over him (e.g., probation officer) on sensitive issues (e.g., drug use) in open-ended fashion without expectation of verification. This is a demand characteristics issue. his combines format or structure of the measure, demand characteristics of the situation in which the measure is taken, and the ego involvement of the provider of the measure. This is not a measure of the extent to which one's behavior is changeable but the changeability of the report of that behavior. Objective measures should rate in the 1-3 range with arrest records for violent crimes=1 and those for status offenses =2. Self-report or a rating by those who are ego involved in some way would be in the 6-7 range. In descending order of ego involvement are: the target juveniles, parents, therapists, teachers, non-blind researchers, CJ personnel. In descending order of response format sensitivity to bias are: self-report, rating, objective count, and independent cross-checking or review.

Rate this measure's potential for social desirability response bias:

Very Low Potential 1 2 3 4 5 6 7 Very High Potential

Confidence in above 2 ratings: [DN7]

Very Low	Low	Moderate	High	Very High
1	2	3	4	5
(Guess)	(Informed Guess)	(Weak Inference)	(Strong Inference)	(Explicitly Stated)

NonDelinquency Variables

Code a separate screen for each nondelinquency outcome measure for which the aggregate treatment and comparison groups can be compared on the first wave of post-treatment outcome. (Subsequent waves and breakouts for this aggregate comparison are coded on separate attachments to be appended to this sheet). Nondelinquency outcome measures are all those that remain after any delinquency outcome measures are coded on the “delinquency behavior outcome measures codesheet” according to the definitions on that codesheet.

Type of construct represented: [N1] Construct represented by this measure (check best one): [Note: Some categories, like “attitudes” occur in various sets below.

Approach this item by first identifying the most appropriate molar category, e.g., psychological adjustment, interpersonal, etc., then finding the best item within that category for the particular measure at issue.]

Psychological adjustment

1 attitudes re delinquency, personal conduct, police, etc.
2 self-esteem, self concept
3 other personality trait
4 behavioral problems checklist, etc.
5 knowledge re drugs, ethics, moral dilemmas, law, etc.
6 mood, anxiety, depression, emotionality, etc.
7 other: _____

Interpersonal adjustment

8 attitudes re interpersonal issues, family, peers, etc.
9 family functioning, communication, household chores, etc.
10 peer relations, etc.
11 social skills
12 other: _____

Community adjustment

13 attitudes re community, citizenship, etc.
14 perceptions by merchants, community officials etc.
15 other: _____

School adjustment

16 attitudes re school, teachers, etc.
17 noncriminal/non-delinquent disciplinary
18 attendance; tardiness
19 dropping out; graduating
20 other: _____

Academic improvement

21 achievement (content mastery in topic area)
22 grades
23 cognitive, general (e.g. IQ)
24 other: _____

Vocational adjustment

25 attitudes toward work, employment, careers, etc.
26 Job attendance, tardiness
27 employment status (gets/keeps job)
28 employment learning (job content, skills)
29 vocational learning (job finding, interview, skills, simulations)
30 other: _____

Adjustment to treatment

31 attitudes re treatment, therapist, program, etc.
32 attendance, participation in treatment
33 treatment progress, e.g., rating
34 status at termination of treatment
35 post-treatment prognosis
36 other: _____

Institutional adjustment

37 attitudes re institution, staff, etc.
38 program behavior, general
39 rule compliance (non criminal)
40 getting along with staff, peers
41 post release prognosis
42 other: _____
43 global adjustment/improvement; individualized criteria (e.g., global rating)
44 all other: _____

Confidence in construct: [N2] Confidence in identification of construct represented by measure:

Very Low	Low	Moderate	High	Very High
1	2	3	4	5
(Guess)	(Informed Guess)	(Weak Inference)	(Strong Inference)	(Explicitly Stated)

Type of measure [N3] (check best one):

1 psychometric/,standardized, multi-item (e.g., achievement, attitude, personality, MMPI)
2 criterion referenced or goal setting; mastery; behavioral objectives--test, form, or questionnaire
3 behavioral observation; behavioral report; behavioral record or charts
4 survey type items, questionnaire, self report form
5 judgment ratings; judgment coding from observation by other(s)
6 archival report (e.g., school, agency records)
7 projective test (e.g., TAT, Rorschach)
8 other: _____
9 cannot tell

Origin of measure [N4](check best one):

1 "off the shelf" named measure or scale
2 taken intact from other research, not in general use
3 adapted or modified from other source
4 pre-existing records or archives
5 new instrument apparently developed for this evaluation
6 other: _____
7 cannot tell

Source of information: [N5] Primary source of information for measure (check best one): [Note: Issue here is who is forming the content recorded in the measure. E.g., if a person fills cut a form or responds to an interview, that person is the information source. If an observer rates or judges another person, however, it is the observer not the person observed, who is the source.]

1 juveniles themselves (e.g., self report, survey)
2 front line service provider; therapist; caseworker
3 program manager, administrator, agency staff, etc. (not front line)
4 researchers acting directly as observers, raters, etc.
5 other observers or participants (e.g., client families, employers)
6 records, archives
7 other: _____
8 cannot tell

Properties of this measure demonstrated, reported, or cited (check all that apply):

Properties demonstrated, validity: [DN1]
Properties demonstrated, reliability: [DN2]
Reliability coefficient: [DN2R]magnitude of coefficient, if given (-99 if missing)
Properties demonstrated, sensitivity: [DN3]
sensitivity/responsiveness/discriminant ability [i.e., indication that measure capable of responding to treatment effect]

Properties demonstrated, none: [DN4] none of the above.

Treatment-test overlap: [DN5] Rate the extent to which the treatment content overlaps or resembles the content of this measure, e.g., as in "teaching the test." At one end of the continuum are measures that are virtual duplicates of the treatment, e.g., a behavioral treatment that reinforces a specific list of behaviors and an outcome measure that counts how often those same behaviors are performed. At the other end of the continuum are measures that have virtually no content similarity to the treatment, e.g., a treatment of insight-oriented counseling about family relations and an outcome measure of math grades in school. This is not a question about the extent to which the treatment caused the dependent variable. The question concerns the content of the treatment not the plausibility of the hypothesized causal relationship. The topic area of the treatment in relation to the topic area of the measure determines the general category. Use the 1-3 range for treatments and measures of generally different content and involving different activities; use 3-5 for those situations like general counseling and delinquency measures where discussion of delinquency may well have been part of the treatment content, giving topic overlap, but the activities of treatment (talking about delinquency) are different from those in the measure (committing delinquency). Use the 5-7 range for fairly clear overlap in both topic area and activity, e.g. substance abuse treatment involving role playing resistance to peer pressure and actual substance abuse incidents as an outcome measure. Within these ranges, adjust for the degree of overlap according to the specifies
of the individual case.

Rate this measure for treatment-test content overlap:[DN5]

Very Low Overlap 1 2 3 4 5 6 7 Very High Overlap

Social desirability bias: [DN6] Rate the extent to which this measure seems susceptible to a social desirability response bias, that is, the extent to which the respondents are (a) able to recognize what response "looks good," (b) may be motivated to "look good," and (a) are able to exaggerate the response in the direction of "looking good." Note that you are not to rate how much social desirability bias you think actually occurred, only how susceptible you think the measure might be. At one end of the continuum would be measures based on objective procedures administered by impartial others, e.g., random surprise urinalysis for drug testing. At the other end of the continuum would be the juvenile's own reports made to someone with authority over him (e.g., probation officer) on sensitive issues (e.g., drug use) in open-ended fashion without expectation of verification. This is a demand characteristics issue. his combines format or structure of the measure,

demand characteristics of the situation in which the measure is taken, and the ego involvement of the provider of the measure. This is not a measure of the extent to which one's behavior is changeable but the changeability of the report of that behavior. Objective measures should rate in the 1-3 range with arrest records for violent crimes=1 and those for status offenses =2. Self-report or a rating by those who are ego involved in some way would be in the 6-7 range. In descending order of ego involvement are: the target juveniles, parents, therapists, teachers, non-blind researchers, CJ personnel. In descending order of response format sensitivity to bias are: self-report, rating, objective count, and independent cross-checking or review.

Rate this measure's potential for social desirability response bias:

Very Low 1 2 3 4 5 6 7 Very High

Potential Potential

Confidence in above 2 ratings: [DN7]

Very Low	Low	Moderate	High	Very High
1	2	3	4	5
(Guess)	(Informed Guess)	(Weak Inference)	(Strong Inference)	(Explicitly Stated)

Effect Size Calculation (ES)

Weeks Delinquency Counted [ES20] (leave blank if nondelinquency variable)

|__|__|__| Approximate (or exact) time period covered by delinquency measure, i.e., period over which counted delinquency occurs, e.g., whether arrested during last six months. (Code number of weeks, rounded to nearest whole number; divide days by 7 and round; multiply months by 4.3 and round; code 999 if cannot tell or NA, but try to make an estimate if possible. Code 888 if total prior history covered).

Weeks Post-Treatment Measured [Time1]

|__|__|__| Approximate (or exact) weeks after end of treatment when measure taken, i.e., what was the interval from the end of the treatment to the time when this outcome measure was taken. (Code whole number, no decimals; divide days by 7 and round to whole number; multiply months by 4.3 and round; code 999 if cannot

tell, but try to make an estimate if possible). [NOTE: If measure was taken more or less immediately at the end of treatment, code this as one week.]

Effect Size Statistics

[Note: Complete as much of this item as possible even if it requires some calculation or manipulation of data presented in the report. Use separate treatment vs. control group statistics if available, otherwise statistics for pooled groups if they are available. If neither available, enter missing data codes.]

Original N

Number of subjects originally assigned/selected for the treatment and control groups before any attrition, dropouts, refusals to participate, etc. (missing=9999). [Note: The issue here is attrition between assignment/selection for treatment and measurement. If attrition data after pretest and after group assignment conflict, code the latter. The three common ways to get information on the original group size are from assignment to treatment groups, the actual pretest data for measures (if there are differences in n between the various pretests, use the largest one) and demographics at pretest. The largest number claimed for each group by any of these sources should be considered the n at assignment.]

_____ treatment group [ES36]
_____ control group [ES37]
_____ total/difference [ES38]
_____ effect size total N if treatment or control N's not known [ES3 by hand]

Effect Size N: Number of subjects whose data is actually represented in the statistics for the outcome on which the effect size calculation is based (missing=9999).

_____ treatment group [ES1]
_____ control group [ES2]
_____ total/difference [ES3]

Effect size total N if treatment and control group Ns not known [ES3]
Mean on measure (missing=999.99)

_____ treatment group [ES9]
_____ control group [ES10]
_____ total/difference [ES11]

Variance on measure (missing=999.99)

_____ treatment group [ES12]
_____ control group [ES13]
_____ total/difference [ES14]

SD (standard deviation) [ES25] [ES26]
SE (standard error) [ES27] [ES28]
Proportion successful [ES29] [ES30]
N successful [ES31] [ES32]
Enter here the raw values for "N Successful" if they are provided. Do not calculate "N successful" from the effect size N and the proportion. Only enter N successful if it is given explicitly.

t-value [ES33]
F-value (df=1) [ES34]
Chi-square (df=1) [ES35]
Enter values as appropriate and available. Note: if you have, or can determine, the proportion or frequency who “failed” or “succeeded” be sure to enter that information.

Effect size (by FileMaker or by hand)
|__|__|.|__|__| ES (two decimals with an algebraic sign in front, plus if favors treatment (i.e., more “success” for treatment group than control), minus if favors control, +9.99 if NA.

Pre-test, Post-test, or Follow-up [ES24]

Identify the type of effect size in terms of the time of measurement of the data on which the treatment vs. control comparison represented in the effect size is made. [NOTE: Code the available information for any dependent variable for which the direction of the difference can be determined (whether favors treatment, control, or neither) even if a numerical effect size value cannot be determined.]
“Pretest” refers to measures of status before treatment or at the beginning of treatment on the same variable used as an outcome measure. E.g., delinquency index for an interval prior to treatment is the “pretest” for the delinquency index for the same length interval subsequent to treatment.
“Posttest” refers to measures of status on first wave of measurement after the treatment is completed.
“Follow-up” refers to measures of status at any wave of measurement after the posttest, i.e., for there to be a follow-up, there must be at least two waves of measurement after treatment is completed; the first would be the posttest, the second (and any others thereafter) would be a followup.

Type of means [ES15] [Note: If ES based on proportion or N successful, code as proportion mean.]
1 arithmetic mean of scores
2 median of scores
3 proportion or rate
4 other: _____
5 cannot tell

Type of variances [ES16]

[Note: If ES based on proportion or N successful, code as proportion variance.]

1 standard deviation
2 variance
3 standard error
4 proportion
5 other: _____
6 cannot tell

Direction of Difference [ES17]

Numerically comparing treatment group scores to control group scores on this measure, the raw treatment vs. control group difference favors (i.e., shows more "success" for) which group (check best one). [Note: Report this information if available even if the numerical values on the variables are not reported.]

1 treatment
2 control
3 neither (exactly equal)
4 cannot tell or statistically insignificant report only

Type of Statistical Test for T-C difference [ES18]

1 no test done
2 kind of test not reported
3 t, F, Z, or r (parametric, no partialling or variance adjustment)
4 Chi-square test
5 other nonparametric test, e.g., Mann-Whitney U
6 test adjusts for covariate, not pretest (e.g., ANCOVA, covariate blocking)
7 test adjusts for PRETEST (e.g., ANCOVA with pretest covariate, repeated measures design, t-test using gain scores)
8 other
9 missing

Statistical Significance Difference [ES19]

[Note: report what the author claims at whatever alpha level, etc. used; if only p-values provided with no statement of what is judged statistically significant, code anything with p<.05 as significant.]

1 significant
2 not significant
3 not reported

Effect Size Confidence [ES22](Confidence in effect size value)

Highly Estimated	Moderately Estimation	Some Estimation	Slight Estimation	No Estimation
1	2	3	4	5

[Note: Confidence guidelines:

5 No Estimation--have descriptive data; can calculate ES directly.

4 Slight Estimation--significance testing statistics rather than descriptive statistics, but have complete stat conventional sort.

3 Some Estimation--have unconventional statistics and must convert to equivalent t-values or have conventional statistics but incomplete, e.g., exact p level only.

2 Moderate Estimation--have complex but relatively complete stats, e.g., multiple regression, LISREL, multifactor ANOVA etc. as basis for estimation.

1 Highly Estimated--have N and crude p value only, e.g., p<.10, and must reconstruct

Page Number Where ES found: _____
Report in which ES found: _____

Appendix 2: DuBois et al. Coding

id 1 2	Study ID
	Report Identification Title
3 4 5 6 7 8 aut	Author(s): (enter first six letters of first author)
9 10 11 12 13 14 jour	Journal (Enter abbreviation of journal title, e.g., JCCP) or ED #
yr 15 16	Year: C.17=Blank
pub 18	Publication Vehicle: Journal Dissertation Book Thesis Paper presentation Govn=t report Private evaluation
sou 19	Source: PsychINFO ERIC Medline Dissertation Abstracts Other Data Base (specify _____) Ancestry (specify Study ID#_____) Research Known to First Author C.20=Blank
prgid 21 22	**Mentoring Program Information and Study Design** Program ID
intgrp 23	Nature of Intervention Group Mentoring alone Mentoring and other intervention (specify_____ _____)
nintgrp 24	Nature of Comparison Group Did not receive an intervention Received mentoring Received intervention other than mentoring (specify_____ _____)

		Setting Where Mentoring Activities Occurred
setting	25	Community School Workplace 4. Institution/Agency/Organization (other than school) 5. Other (specify_____) Unspecified
		Location of Program (City Size)
prgloc	26	Large Urban Small Urban Suburban Rural Mixed
		Program Type
type	27	1. Instrumental (specify_____) Psychosocial Combination Other (specify_____) Unspecified
		Type of Instrumental Focus (if applicable)
instype	28	Educational Employment Other (specify_____)
		Mentor Compensation
comp	29	Educational (course credit, class assignment, etc.) Financial Other (specify_____) None/Volunteer Unspecified
		Mentor/Mentee Match Criteria?
mtcrit	30	0 = No 1 = Yes 2 = Unspecified
		Mentor/Mentee Match Criteria (if applicable; 0=No, 1=Yes)
mtgen	31	Gender
mtrac	32	Race/Ethnicity
mtint	33	Interests
mtset	34	Setting
mtoth	35	Other (specify_____)
		Mentor Screening?
scrn	36	0 = No 1 = Yes 2 = Unspecified
		Mentor Screening Criteria (if applicable; 0=No, 1=Yes)
bkchk	37	Background checks (criminal records, references, etc.)
int	38	In-person interview
hmvst	39	Home visit
		Psychological testing

psytst		40	Other (specify_____)
scroth		41	
train		42	Mentor Training Prior to Match? 0 = No 1 = Yes 2 = Unspecified
amttrn	43	44	Amount of Training (if applicable; code in hours, considering 1 session= 2 hours if only this information is available; round to nearest whole #)
			Characteristics of Training (if applicable; 0=No, 1=Yes)
instrn		45	Instructor-Led
mattrn		46	Prepared Materials Used (e.g., video, workbook, etc.)
indtrn		47	Individual
grptrn		48	Group
ssttrn		49	Self-Study
unstrn		50	Unspecified
super		51	Mentor Supervision? 0=No 1=Yes 2=Unspecified
frqsup	52	53	Frequency of Supervision (if applicable; code # hours per month, considering 1 meeting = 1 hour and rounding to whole#s)
typsup		54	Type of Supervisory Contacts (if applicable) In-Person Telephone Mail Mixed Unspecified
ongtrn		55	Ongoing Mentor Training? 0=No 1=Yes 2=Unspecified
amtont	56	57	Amount of Ongoing Training (if applicable; code # hours per month, considering 1 session=1.5 hours if only this information is available; nearest whole#)
mmcnt		58	Mentor/Mentee Contact Time Expectations/Guidelines? 0=No 1=Yes 2=Unspecified
expcnt	59	60	Expected Frequency Mentor/Mentee Contact (# hours/week, considering each visit to be 2 hours if information is only provided in this form and rounding to whole #s)
		61	Was actual frequency of mentor/mentee contact measured? 0=No

			1=Yes
actcnt	62	63	Actual Average Frequency of Mentor/Mentee Contact (# hours/week considering each visit to be 1.5 hours if information is only provided in this form and rounding to whole #s)
mmlng		64	Mentor/Mentee Length of Relationship Expectations/Guidelines? 0=No 1=Yes 2=Unspecified
explng	65	66	Expected Length of Mentor/Mentee Relationship (# of months, considering 4 weeks=1 month if information only provided in this form, rounding to nearest whole month)
		67	Was actual length of mentor/mentee relationships measured? 0=No 1=Yes
actlng	68	69	Actual Average Length of Mentor/Mentee Relationship (# of months, considering 4 weeks=1 month if information only provided in this form, rounding to nearest whole month)
			END LINE#1/BEGIN LINE#2 (Repeat Study ID C.1-C.2;C.3=Blank)
menage	4	5	What was the average age of the mentors? (round to nearest whole#)
mendev		6	Developmental Level of Mentors Adolescence (12-18 years of age) Early Adulthood (19-29 years of age) Middle Adulthood (30-54 years of age) Late Adulthood (55 and older) Mixed (adult only) Mixed (adolescent and adult) Unspecified
mengen	7	8	Gender of Mentors (percentage male, rounding to whole #)
menwht	9	10	Race/Ethnicity of Mentors (percentages, rounding to whole #s) White/Caucasian
menblk	11	12	Black/African-American
mennta	13	14	Native American
menasi	15	16	Asian-American
menhis	17	18	Hispanic
menotr	19	20	Other int(1)/ext(2)? all (1y/0n) n-mexc.? mon.? par? strucact.? mnsupgrp?
edprlv		22	Educational/Professional Level and Background of Mentors Less than High School Diploma High School Diploma or GED Some College Undergraduate Degree Graduate Degree Mixed Unspecified
			Background in Helping Profession/Role?

helbkg				23	Yes, Occupation/Education Yes, Parent/Caretaker No Mixed Unspecified
ctrl?				24	Is there a control group? 0=No 1=Yes
ctrltype				25	Type of Control: Pretest 2. Random Assignment Static Group
pre?				26	Is there a pretest? 0=No 1=Yes
pretype				27	If there was a pretest, what type is it? Identical to post test Functionally the same as post test
pregrp				28	If there was a pretest, to whom was it given? Given to intervention group only 2. Given to both intervention and control groups
statcov				29	If there was a static group were there non-statistical procedures for creating equality? 0=No 1=Yes
sexcov				30	If the procedure was matching, what variables were used to match (0=No, 1=Yes)? Sex
raccov				31	Race/Ethnicity
agecov				32	Age/Grade Level
schcov				33	School Attended
famscov				34	Family Structure
famicov				35	Family Income Level/SES
achcov				36	Achievement Level
embcov				37	Emotional/Behavioral Adjustment Level
					END LINE#2/BEGIN LINE#3 (Repeat Study ID C.1-C.2;C.3=Blank) **Independent Sample and Deletion Codes**
is	4	5	6	7	Independent Sample Code
delcd				8	Deletion Code (0=Don=t Delete; 1=DeleteBmulticomponent program; 2=DeleteBcomparison within intervention group; 3=Delete--secondary subgrouping of sample)
					C.9=Blank **Participant Characteristics**
					# of Females

nfem			12	13	# of Males
nmal			16	17	# of Both Males and Females (enter only if separate #s for Males and Females not available)
nboth			20	21	Average Age (in years, at start of program, rounded to nearest whole #)
ythage			22	23	Developmental Level Early Childhood (5-8) Middle/Late Childhood (9-11) Early Adolescence (12-14) Middle/Late Adolescence (15-18) Mixed Unspecified
ythdev				24	SES Low 2. Middle High 1. Mixed Unspecified
ythses				25	Race/Ethnicity (percentages, rounding to whole #s) White/Caucasian Black/African-American Native American Asian-American Hispanic Other
ythwht			26	27	
ythblk			28	29	
ythnta			30	31	
ythasi			32	33	
ythhis			34	35	
ythotr			36	37	
risk				38	At-Risk Status Environmental Factors (e.g., single-parent home) Individual Factors (e.g., academic difficulty) Both Environmental and Individual Factors Neither Unspecified ALSO SINGPAR(<=75%) [1Y/0N]:
					END LINE#3/BEGIN LINE#4 (Repeat Study ID C.1-C.2;C.3=Blank) Independent Sample ID **Outcome Variable Information**
crit			4	5	Criterion Measure Self-Esteem/Self-Concept Perceived Self-Efficacy/Sense of Mastery Classroom Behavior Report Card Grades School Attendance Academic Achievement Test Scores Academic Self-Concept/Self-Esteem Attitudes Toward School

School Drop-Out
Intelligence/Cognitive Skills and Abilities
Substance Use
12. Substance Use Attitudes/Knowledge
Problem/High-Risk Behavior (other than Substance Use)
Psychological/Emotional Distress
Psychological/Emotional Well-Being
Peer Relationships
Family Relationships
Social/Cultural Activity
Social Skills/Social Competence
Coping Behavior
Community Service
Other (specify_____)

sour ____ 6

Data Source
1. Youth
Parent
1. Teacher
Mentor
Administrative Records
Other (specifiy

foll ____ 7

Time of Data Collection for Post-Test Relative to End of Program
During mentor relationship
Immediate post-test
Follow-up

folllng ____ 8 ____ 9 ____ 10

If follow-up assessment, what was length of interim period from the end of the program? (specify in weeks, rounded to nearest whole #)

C.11=Blank

Statistical Outcomes

Group 1 (Intervention)

intpre ____ 12 ____ 13 ____ 14 . ____ 15 ____ 16

0 (pre-test)

intsdpr ____ 17 ____ 18 . ____ 19 ____ 20

SD

intpst ____ 21 ____ 22 ____ 23 . ____ 24 ____ 25

0 (post-test)

intsdpst ____ 26 ____ 27 . ____ 28 ____ 29

SD

nint ____ 30 ____ 31 ____ 32 ____ 33

n

C.34=Blank

Group 2 (Control)

cntpre ____ 35 ____ 36 ____ 37 . ____ 38 ____ 39

0 (pre-test)

____ ____ . ____ ____

SD

cntsdpr		40	41	42	43	
cntpst	44	45	46	47	48	0 (post-test)
cntsdps		49	50	51	52	SD
ncnt		53	54	55	56	n
						C.57=Blank
dtype					58	Type of d Post - pre $Post_{int}$ - $post_{cnt}$
dder					59	How was the d index derived? 1Unadjusted means comparison 2ANCOVA/MR w/pre-test only as covariate 3ANCOVA/MR w/only control measure[s] other than pretest 4ANCOVA/MR w/both pre-test and other measures as controls 5Blocking Design
ddir					60	Direction of Effect Size Positive Negative
dsig					61	Significance of Finding (p < .05 two-tailed; 1=yes; 2=no; 3=not available)
d		62	63	64	65	d index or

Appendix 3: Tolan et al. Additional Codes for Mentoring Meta-analysis

1. Differentiate Risk into:
1. Behavioral (aggression, delinquency level, etc.)
2. Environmental-Individual Differences such as Family, School Achievement
3. Hi-Risk Setting such as Violent Community
2. Mentoring Activities Included (yes/no)
1. Emotional Support
2. Teaching/Information Provision
3. Advocacy
4. Modeling
5. Acting as Identification Figure
3. Nature of Relationship/Basis of Mentoring
1. Survivor (had same issues)
2. Civic Duty- as part of job or otherwise volunteer to help those in need
3. Professional Development or Duty
4. Other
4. Implementation Quality
1. Checked or not
2. Fidelity or Application of Key Principles Checked
3. Retention of Participants in Program (percent)
4. Retention of Participants in Study (percent)
5. Evidence that Mentors Retained, yes/no, percent