THE CAMPBELL COLLABORATION

Protocol for a Systematic Review: Policing Schools Strategies to Reduce Crime, Increase Perceptions of Safety, and Improve Learning Outcomes in Primary and Secondary Schools

Anthony Petrosino, Sarah Guckenburg, Trevor Fronius

Submitted to the Coordinating Group of:

☒	Crime and Justice
☒	Education
	☐ Disability
☐	International Development
	☐ Nutrition
☐	Social Welfare
☐	Other:

Plans to co-register:

☐	No		
☐	Yes	☐ Cochrane	☐ Other
☐	Maybe		

Date Submitted: 13 May 2013
Date Revision Submitted: 1 November 2013
Approval Date: 15 January 2014
Publication Date: 03 March 2014

BACKGROUND

The Problem, Condition or Issue

The presence of gangs, drugs, bullying, weapons, and other criminal activity is associated with negative effects on school culture and student learning (e.g., Henry, 2000). Schools and school districts have long turned to police interventions to address these and other challenges to improve school safety (Shaw, 2004). In turn, police departments have also recognized the importance of developing strategies in collaboration with schools, such as special truant units or police-school liaisons, to combat school-based crime and disorder or to develop more positive relationships with youth (Shaw, 2004).

Although school-based police interventions are controversial (see below), American educators and the public are increasingly open to their implementation (Cook, Gottfredson, & Na, 2009). Police presence on school grounds has grown dramatically following several high-profile shootings (Brown, 2006). In particular, the recent massacre at the Sandy Hook Elementary School in December, 2012, has renewed calls for deploying armed police in schools, a proposal offered by both President Obama and the National Rifle Association. Increased police presence and intervention at schools has also been reported in the United Kingdom and other European nations (Smith, 2003).

Research suggests that juvenile violence has been on the decline in the United States since the mid-1990s (Gottesman & Schwarz, 2011), and schools remain safe places for students (relative to their time outside of the school setting). However, a considerable amount of crime and misbehavior has always occurred in educational settings—a result of placing large numbers of children together in one setting. For example, the U.S. Departments of Education and Justice report that during the 2009-2010 school year, 85% of public schools reported that at least one violent crime, theft, or other crime occurred at their school (Robers, Zhang, & Truman, 2012). Over half of administrators surveyed in the 2009-2010 School Survey on Crime and Safety reported at least one student threat to physically attack another person, with or without a weapon (Neiman, 2011). Moreover, in 2007, 8% of students in high school reported being threatened or injured with a weapon during the previous 12 months (Dinkes, Kemp, & Baum, 2009).

The problem of school violence, crime, and disorder is not just an American issue. Estimates of school-related crime from other countries are more difficult to obtain, but the European Observatory on School Violence (EOSV) has been collating data from nations such as the U.K., Germany, France, and Spain (Debarbieux, 2003). For example, the EOSV has found that nearly 6% of U.K. teachers report being threatened or attacked at school, 20% of Spanish students report that they have committed some form of vandalism at school, between 15-46% of German students report carrying a weapon to school, and over 25% of French students report being involved in a physical assault on school property (Debarbieux, 2003).

These international estimates dovetail with a recent World Health Organization (WHO) report that listed violence as the world's priority public health issue (Krug, Mercy, Dahlberg, & Ziwi, 2002). The WHO report has generated a renewed focus on youth violence and, more specifically, on school violence, in such organizations as the United Nations Economic, Social and Cultural Organization (UNESCO, 2007). The European Observatory's offspring, the *International Observatory on Violence in the School Environment,* was inspired in part by growing interest in this issue by nations around the globe.

The Intervention

Given the high levels of crime and disorder present at schools in the U.S. and in other nations, it is not a surprise that police and educators would collaborate in order to combat it. Shaw's (2004) international review of police-school practices categorized such strategies into three groups: (1) school-based police officer programs; (2) police as 'teachers' (i.e., curriculum approaches such as the police officer teaching the students about drug, gang, or sexual abuse prevention); and (3) comprehensive or broad-based liaison programs in which police and other social service agencies are involved with the schools. Shaw reported that non-curricular policing programs (e.g., police patrol strategies, problem-oriented or community policing tactics, assigning school resource officers or police-school liaison offices, or other measures) have been implemented in a wide range of countries, including the U.S., U.K., Canada, Germany, Denmark, Australia, the Netherlands, Belgium, Estonia, Poland, and South Africa, at least through 2003. Shaw's review was globally comprehensive but not systematic; she concluded her synthesis (2004, p. 26) with a future research agenda that includes "a more comprehensive, comparative review of police-school interventions" and "the systematic collection and documentation of good practice models and tools."

Why it is Important to do the Review

To our knowledge, there has not previously been a systematic review of evaluative evidence specific to the "policing schools" area, with the exception of police-led curriculum efforts such as Drug Abuse Resistance Education (D.A.R.E) (Ennett, Tobler, Ringwalt, & Flewelling, 1994; West & O'Neal, 2004). In other words, while police departments maintain a significant presence in many schools, there is not yet a research-based consensus on whether School Resource Officer or special police-school liaison programs, or other policing schools strategies make schools safer and reduce criminal or delinquent behavior and non-criminal student misbehavior (such as bullying and harassment). There is also no agreement on whether the presence of police enhances perceptions of safety by students or staff.

Critics of school-based policing argue that police in schools represent an unnecessary expense and may have unintended negative consequences for students, teachers, and staff (e.g., Petteruti, 2011). For example, police presence in the schools could lead to behavior normally handled informally by schools to be criminalized by arrest and formal charges (e.g., Petteruti, 2011). The National Association of School Psychologists (2013) recently issued a brief that summarized a number of potential harms from increasing police and security at

schools, including sending a signal to students that their campus is unsafe, increasing fear, and undermining school climate.

In 2012, Petrosino, Guckenburg and Fronius reported on a scoping study using systematic review techniques to assess the state of the evaluation evidence in this area. They located 11 quasi-experiments available that had a comparison group. Only one would have met the minimum evidentiary standards of the What Works Clearinghouse in education, and most of the studies would have been scored a "3" by the Maryland Scale of Scientific Methods: a common approach to rating evaluations in justice. However, they did not attempt any meta-analysis, and their review covered the literature through 2009.

This proposed systematic search is designed to build upon Shaw's (2004) work and recommendations, and the earlier Petrosino, Guckenburg and Fronius (2012) scoping study, to provide a systematic review of research on the effects of policing schools strategies.

How the Intervention Might Work

A general theory of change for policing schools strategies is provided in the Figure in Appendix A. As the Figure denotes, there are a number of anticipated interventions that we expect fall under the "policing schools" strategies, including various police patrol strategies, school resource officer allocation, police-school liaisons, safe routes/corridor programs, and community and problem-oriented policing strategies specifically targeting the schools. Programs that solely involve the police officer teaching a specific prevention curriculum such as D.A.R.E., G.R.E.A.T., or sexual abuse prevention are not eligible, though interventions that combine curricular and non-curricular strategies would be included. Policing programs that involve providing information on school crime or data to schools would be included, if the police are not teaching a curriculum or lesson plan.

These programs and strategies are expected to increase police presence activities and interaction with students and staff at the schools. Police presence serves as a deterrent to crime and misbehavior, and police establish relationships that allow them to gather information to prevent or intervene with students who may be planning a crime. In turn, this should reduce criminal behavior and non-criminal student misconduct; increase perceptions of safety by students, staff, and parents; and lead to more positive relationships between police with young persons and with the schools. This should also result in increased attendance, test scores, and grades, and improved perceptions of school climate. However, there is the possibility of unintended negative consequences that police presence leads to heavy surveillance, resulting in police coming across more misbehavior by youth, leading to overuse of arrest for behavior normally handled informally by schools and to more push-outs of students (disproportionately affecting racial and ethnic minorities, and youth with disabilities). Police presence could also lead to increased fear of crime and misbehavior by sending a signal to the school community that the building is unsafe.

The types of learning outcomes we anticipate are grades, attendance/absenteeism, dropout,

test scores, and perceptions of school climate. The types of non-educational outcomes we anticipate are measures of safety such as police reports, calls for service, arrests, school disciplinary reports, self-reports of student behavior, self-reports by staff of safety and fear, self-reports by students of safety and fear, and measures of parent fear for their youth at school.

OBJECTIVES

The objectives are to systematically review the research to answer three research questions relevant to potential positive effects of police-school programs, and two questions relevant to a potential unintended negative consequence.

Potentially positive effects

1. What are the effects of policing strategies in primary and secondary schools (K-12 in USA) on measures of crime and non-criminal student misbehavior (e.g., certain levels of bullying, harassment, etc.) in and around schools?

2. What are the effects of policing strategies in primary and secondary schools on measures of staff or student perceptions of safety, climate, and culture?

3. What are the effects of policing strategies in primary and secondary schools on learning outcomes (e.g., test scores, grades, attendance, etc.)?

Potentially negative, unintended consequences

4. What are the effects of policing strategies in primary and secondary schools on arrests, formal sanctions, or push-outs of students?

5. Were these effects disproportionate to racial and ethnic minorities or students with disabilities?

METHODOLOGY

Criteria for including and excluding studies

For this project, we will only include those studies that have the following characteristics:

1. *The study must evaluate a policing strategy whose primary approach does not involve police officers teaching a curriculum*. The program must focus on a goal such as reduced crime or increased school safety. This would include strategies such as increasing the number of police patrols on school grounds. Police must either lead or be a strong partner in the strategy. Broader programs that involve multiple agencies with minimal police involvement (e.g., police just attend a meeting of concerned agencies about a school problem) will not be included. We will not include

evaluations of programs that include the police along with many other components (e.g., police collaborations with health and other social service agencies). In such studies, one cannot isolate the effects of the police action. The review will not include studies of the impact of academy or educational training on police performance.

   a. There have been meta-analyses of police-taught curricula such as D.A.R.E. (e.g., see Ennett et al., 1989). The role of the officer (as a teacher) is very different in such strategies in that the focus is to deliver universal prevention strategies to students, not to gain immediate crime prevention and safety benefits at the schools. Indeed, there have been over a dozen high-quality studies of D.A.R.E., but not one collected crime outcomes; instead, they focused on knowledge about drugs, attitudes and perceptions of youth, and self-reported drug use (Petrosino, 2005). Thus, evaluations of police-led education programs such as D.A.R.E., G.R.E.A.T., or sexual abuse prevention curricula will not be included. However, programs that involve the combination of a curriculum component with a non-curricular policing strategy will be included.

2. *The policing program must specifically focus on crime prevention and safety of a primary or secondary school (Kindergarten to 12th grade in U.S. school systems).* The strategy must specifically target schools; projects that include schools in a wider community or neighborhood intervention are not eligible for inclusion in the review. Studies that test programs implemented within primary or secondary school grounds, or are focused specifically on its students (for example, providing safe passage to school, adding school resource officers, etc.), will be included. We will not include studies of policing programs implemented within preschool or higher education settings. The implications of policing preschool facilities and college/university settings are different than for policing primary/secondary schools. For example, most colleges and universities, at least in the United States, have their own police forces. In addition, students at the higher education level are adults, and the campus setting involves a less controlled environment than typical K-12 settings. In our earlier review searches, we did not come across a single study examining a police program at a preschool; this makes sense given the young age and lack of criminal intent of the population. However, elementary schools, although they also have a very young population, often are the site for school resource officer programs (e.g., Jurney, 2013).

3. *To be included in estimates of effect, the evaluation must include a distinct randomized control or quasi-experimental comparison group, or employ a time-series analysis (if only a single group was studied).* There are many types of quasi-experiments (e.g., see Shadish, Cook, & Campbell, 2002), but our focus is on those quasi-experiments that either included a comparison group or, if studying a single group, the study must have used a more sophisticated interrupted time series (ITS)

design, with a minimum of 20 time intervals in the time series (e.g., 10 months before and 10 months after the intervention begins). Such ITS designs are considered to have stronger internal validity than simple pre-post designs (comparing a before and after period). Although pre-post designs are more common, they will not be included because of their susceptibility to internal validity threats (e.g., Mark & Reichardt, 2004). For similar reasons, we also will not include studies that compare results for a single school to a larger pool of schools in any city or state. Lastly, we will exclude studies that used principal survey data to identify correlations between police presence or absence and principals' perceptions of school-based crime.

    a. However, we will collect and report on non-experimental evaluation and correlation studies to provide context and flush out important issues relevant to context, theory, implementation, and policy.

4. *The evaluation must include at least one outcome measure that reflects crime, misbehavior, perceptions of safety, or student learning.* Measures may include such data as official police reports, arrests, calls for service, school disciplinary records, student self-reports of victimization or delinquency, truancy, staff perceptions of school safety, drop out, test scores, attendance, and grades.

5. *The study report is available through 2013, without regard to document type (published or unpublished), nation of study origin, or language.* We will search for studies published or available up to and including 2013, without regard for the start date of publication. In concert with Campbell principles, we will target unpublished (grey literature) and published studies, and also attempt to find English and non-English studies without regard to the nation in which the study was conducted.

6. *The study can include samples of students, teachers/staff, individual schools, or school districts as the unit of analysis.* Given that we desire to construct as broad a search as possible to identify evaluative studies, we impose no exclusion or inclusion criteria on the basis of the type of sample in the study. Programs may target schools, districts, students, or professional staff, and samples may be comprised of individuals or larger aggregate units.

Example of a study that would be included in our review

A quasi-experimental evaluation that would meet the study inclusion criteria was reported in the United Kingdom (Bhabra, Hill, & Ghate, 2004). The study evaluated the location of one police officer in each of 11 schools located within ten designated “crime hot spots” in the United Kingdom. The results were compared to two schools that were also in high crime areas that did not have a police officer assigned.

Example of a study that would not be included in our review

Esbensen, Osgood, Taylor, Peterson, & Freng (2001) reported on a quasi-experimental evaluation of a police-taught curriculum designed to prevent adolescent gang involvement, Gang Resistance Education Awareness Training (G.R.E.A.T.). Because this study involves police teaching a curriculum in the schools, it would not be included in this review.

Search strategy for identification of relevant studies

We will rely on five major strategies to identify eligible evaluations published through 2013. These are:

1. *Electronic searches of bibliographic databases*. Researchers will use available online resources and databases at WestEd, the University of Pennsylvania, George Mason University, and Bridgewater State University, including *Criminal Justice Abstracts*, the *National Criminal Justice Reference Service* (*NCJRS*), *Sociological Abstracts*, and the *Education Resource Information Center* (*ERIC)*. See Appendix B for a listing of electronic databases that will be searched.

2. *Checking the bibliographies in prior systematic and narrative reviews*. Besides the aforementioned Shaw (2004) review, there are systematic and narrative reviews of related topics that may have captured studies eligible for inclusion in this project. For example, reviews of research on the effects of strategies to reduce school violence (e.g., Mytton et al., 2006; Derzon & Wilson, 1999) could include police-involved strategies if they met other criteria for those reviews. We will identify the reviews (via such sources as the Campbell and Cochrane Libraries and U.K. Centre on Reviews and Dissemination's Database of Reviews of Effectiveness), retrieve those documents, and inspect the citations in these reviews.

3. *Google searches of the Internet*. Many institutions are putting their evaluation reports on the World Wide Web. A well-crafted Google search strategy can uncover some evaluation reports that are not indexed into the bibliographic databases and websites mentioned above. We will also identify selected think tank, professional association, and government websites for targeted searches (e.g., National School Safety Center, National Association of School Resource Officers, the Virginia Center for School Safety).

4. *Citation chasing*. The reference section of every retrieved evaluation report will also be checked to determine whether any possible eligible evaluations are listed. As noted in the eligibility criteria, we are not exclusively seeking English language reports. There are also a number of general articles on policing and schools (e.g., Shaw, 2004) that reference evaluative studies, and any relevant citations from these papers will be checked.

5. *Contacting the "informal college" of researchers on this topic and professional associations of practitioners*. There is an active network of researchers conducting or

aware of evaluative studies on this topic. We will reach out to such researchers by email to query them about studies that may be relevant to our project. We will also contact professional associations of practitioners in education and juvenile justice, including the International Association of Chiefs of Police (IACP) and the National Association of School Resource Officers (NASRO). All individuals and agencies contacted will be listed, with results of requests. We will ask our colleagues from other nations for help in identifying any non-English studies. WestEd also has employees bilingual in Spanish, French, Japanese, and Chinese who can translate abstracts or full-text documents in non-English to determine their eligibility for this search project.

Keyword strategies for bibliographic databases

The databases listed in Appendix B can be somewhat idiosyncratic. Our approach is to conduct pilot searches of terms, working iteratively until the yield of citations and abstracts is as relevant as possible. In other words, we want to maximize sensitivity (getting as many citations and abstracts as possible) and specificity (making sure that as many as possible are relevant to the project).

Our planned search strategy combines three types of keywords. The first set of keywords (and their derivatives) targets outcome studies: e.g., “random,” “experiment,” “control,” “evaluate,” “trial,” “impact,” “effect,” and “outcome.” A second set of keywords focuses the search on schools and education institutions: e.g., “student,” “school,” “district,” “classroom,” “academy,” “,” “campus,” “teacher,” “principal,” “faculty,” “bully,” “truancy,” and “superintendent.” The third set of keywords focuses the search on policing, including keywords such as “police,” “patrol,” “sheriff,” “constable,” “enforcement,” “officer,” and “security.” When the database does not permit extensive lists of keywords to be combined, simpler searches involving words such as “police” and “schools” will be used.

For example, a proposed search in Criminal Justice Abstracts would be:

Query: AB=(Columbine or school* or student* or campus* or teacher* or truan* or vandal* or classroom*) and AB=(experiment* or evaluat* or assess* or impact* or outcome* or effect* or randomly or randomize* or "comparison group" or "control group" or controls or comparisons or "control condition" or "comparison condition" or “time series”) and AB=(police* or "law enforcement" or patrol* or policing or "security guard" or "security measure*" or "crime prevention" or sheriff* or "cop" or "cops" or constabl* or detective* or undercover or "school crime" or "school violence" or "school safety" or "school security" or "resource officer*) or “problem-oriented”).

Retrieving and final screening of studies

Search methods will usually result in a large number of citations and abstracts. The review author conducting the search will save the results from the searches described earlier and

mark those for which full-text should be retrieved. This will be examined by at least one other co-author to ensure that there is agreement on the selections. Since all three authors are experienced in conducting reviews of this type, no training is necessary.

Many of these will be easily excluded as not being relevant to the proposed review. In some cases, however, they will identify potentially eligible studies. The full text documents of those potentially eligible studies will be retrieved and screened by the co-authors before the study can be formally included. We expect to retrieve the full text of most articles electronically. We also have access to interlibrary loan through Bridgewater State College. When a full text report is received, two of the review co-authors will read it to ensure that it meets the eligibility criteria. If there is no agreement on the inclusion of a particular study, it will be excluded and the exclusion documented in the final report.

We will also establish a bibliographic reference database to maintain a log of all included and excluded studies. The log includes a field that allows the research team to document the reason for exclusion.

Extracting information from each study

We have designed a preliminary instrument to guide us in recording information from each study (see Appendix C). Although the instrument contains several open-ended items, these will be collapsed when appropriate into a smaller number of categories to permit further analysis. For example, items such as "how equating was performed" can be collapsed into three or four larger categories representing the most frequent responses (e.g., discontinuity, covariate matching, propensity score, post-hoc statistical matching) and an "other" category that captures all those responses that do not fit into the most common methods of equating in this set of studies. The instrument has items in the following areas:

Researcher, study, and contextual characteristics

Study reports can be used to provide information about the publication and characteristics about the study and its context. For example, we will extract data about the type of publication and the setting in which the study was conducted. If the documents provide more detail on the setting and political/social context in which the study takes place, we will also include it.

Study methods and methodological quality

We will extract information about the randomization, quasi-experimental assignment, and other methodological aspects of the evaluation. The level of assignment and whether the study included multiple analyses at different levels will also be coded. It is especially critical that information about three key issues in the implementation be extracted from each study report:

1. *How the groups were equated and whether any problems with equating were*

reported. The integrity of a randomized experiment or a quasi-experiment largely rests on how faithfully the equating procedures were implemented. We will code information about randomization and the quasi-experimental matching or equating procedures that were used in the study. In randomized experiments, this includes how much of the originally randomly assigned sample actually received the treatment (slippage from the "intention to treat" sample). We will code this information using a two-stage process. The first stage is a more detailed gathering of the facts about the assignment. The second stage will be comprised of ratings by two reviewers that will indicate the degree to which group equating was compromised by any reported problems.

2. *Whether the researchers report a loss of participants from the initial assigned sample at the end of the study, how much attrition is reported, and whether the attrition differentially affects one group or the other*. Such attrition, if it is significant, can comprise the equating of groups, particularly if different types of people drop out from the intervention than dropped out from the other conditions. We will code specific information on the amount of attrition (if it occurred) and whether it was differential in nature. We will conduct sensitivity analyses to examine whether results change when we exclude studies that experienced either different levels of overall attrition (10%-25%; 25%-50%; 50% or more) or differential attrition within studies (significant losses from one group or the other, with a difference in attrition of 5% or more between the groups). These are obviously subjective classifications, but the goal is to determine if the attrition compromised the study findings.

3. *Whether the program experienced significant implementation and fidelity problems*. The first two issues deal with the implementation of the evaluation. This issue deals with the implementation of the program; there may be no observable program impact because no "real program" was ever implemented. We propose two-stage coding of implementation. First, we will code, in descriptive and qualitative form, any implementation problems noted by the investigators. Second, we will then rate the degree of implementation problems (with the standard being how the implementation problem affects a "fair test" of the program under investigation) as "high," "moderate," or "low."

Intervention and control conditions data

These items will solicit detailed descriptions of the intervention and control conditions, including the "dosage" of the treatment being implemented, and the number of participants assigned to each group. The types of interventions that we expect include School Resource Officer Programs, police-school liaison programs, problem-oriented policing strategies, and police patrol/surveillance at the school to deter crime within or around the school, or to ensure safe routes for children as they walk to school. We anticipate that the evaluations in

this review sample will be comprised of a single intervention and a single control group. When this is not the case, we will select the most policy-relevant groups to compute our experimental versus control condition contrast. In most cases, it will be the groups that experience the greatest contrast between conditions, i.e., the most intensive intervention condition versus the least intensive control condition. We recognize the importance of documenting these decisions for full transparency.

Participants in the study

These items solicit detail about the type of participants in the trials, including information on the country where the study took place, the nationality of the participants, the age and school level targeted, gender, whether an urban or rural setting was involved, and the socioeconomic status of the students.

Outcomes

For each eligible study (each study will have, at minimum, one outcome measure meeting the eligibility criteria described earlier), we will extract information on reported outcomes of crime (e.g., specific delinquent or criminal acts, arrests, calls for service, referrals to police, disciplinaries, etc.) and other behavior (bullying, aggression, etc. that does not rise to "criminal" act but represents misbehavior), perceptions of safety/fear, and learning outcomes. We will also code any other outputs or data on key "mechanisms" that would provide clues as to why the intervention did or did not have its intended impact.

Handling multiple reports on the same experiment

Note that investigators may publish several articles on the same study. Our unit of analysis is the individual evaluation and not the individual research article, and so it is reasonable to extract information from all documents to complete the coding instrument for one experiment. When reports on the same study contain conflicting information, we will employ a number of strategies, including contacting the original investigator(s) for resolution.

Criteria for determination of independent findings

Each study will be represented by a *single effect size* per outcome to prevent the analysis from being compromised by non-independence (multiple effect sizes from one study). Although some evaluations may report just a single outcome at one time interval, it is possible that other evaluation reports will include analyses at various time intervals and may use various constructs that reflect school safety and learning outcomes. Therefore, decisions have to be made about what outcome will represent the effect size for that study.

For this review, we will keep the outcomes distinct. That is, we will analyze crime/behavior, perceptions, learning outcomes, and potential unintended negative consequences separately. We do not know as of yet how such outcomes will be reported: i.e., will they be prevalence measures (percentage of groups that enroll or attend) or incidence measures (the mean rate

for some outcome of interest, such as the mean number of discipline reports per student). If results are varied and include prevalence and incidence rates, we will discuss the best way to report these (combine and average, separate out, etc.) and make such decisions explicit in our review. We expect to find a range of measures used to assess each of these four broad types of outcomes, and we plan to organize these measures into useful constructs. For example, within learning outcomes, achievement and attendance may be common categories.

Although we expect that many reports will only include an analysis at a single point in time, we propose to report two different analyses to handle the studies that report outcome data at various time intervals. We will report effect sizes at first follow-up (the first time interval reported) and the longest effect (the effect size for the longest follow-up period). If one time interval (e.g., 1 year) is reported in the study, it will be used in both analyses.

If regression-adjusted estimates are reported for the experimental versus control groups, we will rely on them for any quantitative synthesis since they theoretically reduce statistical "noise" that may have come from chance fluctuations or randomization violations (in the case of well implemented experiments) or uncontrolled variables (in the case of quasi-experiments).

Some studies report analyses at multiple levels (e.g., Schultz, 2004), i.e., for schools or localities and for students. Our rule is to prioritize capturing this information at the individual student/staff level. If randomization or assignment is done at the group level (i.e., school or district), we will adjust standard errors to take clustering into account; this was a procedure we used in a prior meta-analysis for the Campbell Collaboration (see Petrosino, Morgan, Fronius, Tanner-Smith, & Boruch, 2012). Those studies in which it is not possible to adjust for standard errors at the individual level will be analyzed separately.

Description of coding process

To ensure that we achieve good coding reliability, we will have two of the coauthors read and record information from all reports if the yield of studies is less than 30 and from a randomly drawn 50% of the studies if the yield is more than 30. We will assess coding reliability (i.e., inter-rater agreement) by using the percentage of agreement for each item before conflict resolution procedures (see below), rather than reporting a global inter-rater reliability statistic. This will avoid inflating reliability measures with study characteristics that generally achieve perfect agreement (e.g. year of publication) with those that do not. Items with lower rates of agreement (less than 80%) will be investigated to determine the source for conflict. The two main coders (Guckenburg and Fronius) will meet with the lead co-author (Petrosino) to resolve disagreements and discuss coded items. Both will explain their rationale for how they coded, and if there is consensus among the three parties, the item will be retained. We will drop those items from our database in which resolution could not be reached, as well as items that lack clear interpretation.

Statistical procedures and data synthesis

Data will be entered into the Comprehensive Meta-Analysis (CMA), version 2. We will use CMA to statistically combine results from the evaluations.

We will report standardized mean differences (Hedges *g*), which is similar to Cohen's effect size measure, but it includes a small sample bias correction for *d*, for each main effect size analysis. Hedges *g* (like Cohen's *d*) is a very flexible effect size metric, and many formulae are available to estimate it from information often reported in evaluation articles (e.g., statistical test data, probability levels). The effect sizes will be inverse-variance weighted when combined; procedures are available in CMA to do this. Forest plots will be used to display the results from the effect sizes. The plot will display, for each study, the effect size, confidence intervals, and significance level. The plot will also display the same for the average effect across studies. Note that all analyses will be reported assuming a random effects model, and the estimate will be weighted by the inverse variance. When describing results in the text, we will report the effect size, the confidence intervals, and whether the analysis indicates that the result is statistically significant. Note that a positive effect size indicates that the strategy had a beneficial impact on the outcome (e.g., a decrease in crime or an improvement in grades); a negative value effect size indicates a harmful effect for intervention (0 means there is absolutely no difference in values of the outcome measure between treatment and control groups).

Because of the likely heterogeneity in interventions, samples, and settings, we will assume random effects models in our analyses, which tends to be more conservative than the fixed effects approach. For our analyses, we will conduct tests for heterogeneity to determine if the average effect size is a good representation of the sample of studies being used in the analysis. We anticipate the heterogeneity will be present, given the suspected variations in intervention type, sample populations, and the like in these studies.

Using CMA, we will confirm heterogeneity in each summary analysis (of each outcome at each of the two time intervals: first and longest) through the Q-Value, which is reported as a summary indicator of the extent of variance across studies in the sample. We will also report I-squared and tau-squared values as additional checks on heterogeneity. However, moderator analyses have to be approached carefully, as they are often based on small numbers of studies (the "small cell" problem), and such analyses can be significant by chance if large numbers of variables are considered (the "capitalizing on chance" problem). We anticipate examining a small number of moderating variables as a source of heterogeneity, by comparing the effect sizes (assuming a random effects model) for the following potential moderators:

- *Intervention type*. An important policy question is whether communities and schools are getting more "bang for the buck" using one particular approach or another. We will recode the detailed intervention categories into discrete groups for analysis and compare the average effect sizes for these groups.

- *School level*. Because of the different age levels and population differences at different school levels, we will examine effects for primary (elementary schools in the US context) and secondary schools (middle and high school in US context) separately.

- *Research design*. Because a variety of methods may be used to study policing schools strategies, we will examine the difference in effect size for experimental and different types of quasi-experimental design. Prior research on the effects of design choice on program impacts has been mixed; for example, Weisburd, Lum, and Petrosino (2001) did report a sizable difference in conclusions in quasi-experimental studies compared to randomized experiments, while other research indicates trivial difference between experimental and quasi-experimental design (Lipsey, 1992).

- *Publication biases*. We will examine the difference in effects for reports published in peer-review journals or books versus other literature (e.g., dissertation theses, government reports, etc.). We will also include a funnel plot and conduct an Egger regression test for funnel plot asymmetry (1997). Results from the Egger and colleagues' (Egger, Davey Smith, Schneider, & Minder, 1997) test for funnel plot asymmetry will be used to indicate if there is a significant positive association between the effect size and standard error indicating possible evidence of publication bias.

We are open to including other moderating analyses, if appropriate to conduct (i.e., there are enough studies to warrant subgroup analyses), such as looking at the relationship between the characteristics of the schools and outcomes.

Treatment of non-experimental research

As mentioned earlier, we will collect non-experimental evaluations and correlation analyses. Our plan is to summarize them in a separate section, and in particular, to use them to highlight the context, theory, implementation, process, mechanisms, and challenges to policing schools programs.

Treatment of qualitative research

It is very unlikely, because of our focus on experiments and quasi-experiments, and our focus on quantifiable outcomes that can be converted into an effect size metric, that we will uncover much qualitative research. However, we will code the presence or absence of ancillary qualitative studies, what the studies focused on, and what the main findings are. Certainly, qualitative data from the experiments and quasi-experiments will be used to illuminate three particular areas: (1) the context for the intervention; (2) the theory or mechanisms by which the program is supposed to impact the ultimate outcomes; (3) the quality and nature of the intervention and comparison condition; and (4) potential issues that may be important to flush out and discussed in relation to the outcome or moderating

variable analyses.

Treatment of economic data

We will report on any economic data included in the primary studies that are included in the review. Our prior experience in conducting reviews is that cost data are rarely included, but it remains worthwhile to include. This includes information on the costs of the program, any analysis of the cost-effectiveness of the intervention (e.g., the cost per school), and cost-benefit studies (e.g., the sum costs and benefits of the program). It is important that this information be linked in some way to the primary outcome studies so that it can be retrieved during the searches.

REFERENCES

Bhabra, S., Hill, E. & Ghate, D. (2004). *National evaluation of the Safer School Partnerships programme*. London: Youth Justice Board for England and Wales. Retrieved from http://www.yjb.gov.uk/Publications/Resources/Downloads/Safer%20School%20Partnerships%20Evaluation.pdf

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis (Version 2) [Computer Software]*. Englewood, NJ: Biostat.

Brown, B. (2006). Understanding and assessing school police officers: A conceptual and methodological comment. *Journal of Criminal Justice, 34*(6), 591-604. doi: http://dx.doi.org/10.1016/j.jcrimjus.2006.09.013

Cook, P. J., Gottfredson, D. C., & Na, C. (2009) *School crime control and prevention*. (Working paper series). Cambridge, MA: National Bureau of Economic Research.

Debarbieux, E. (2003). *Violence in schools: A European review*. Presented at the Hamilton Fish Institute on School and Community Violence, Washington, DC.

Derzon, J. H., & Wilson, S.J. (1999). *An empirical review of school-based programs to reduce violence*. Presented at the annual meeting of the American Society of Criminology, Toronto, Canada.

Dinkes, R., Kemp, J., & Baum, K. (2009). *Indicators of school crime and safety: 2008* (NCES 2009-022/NCJ 226343). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315,* 629-634. doi: 10.1136/bmj.315.7109.629

Ennett, S. T., Tobler, N. S., Ringwalt, C. L., & Flewelling, R. L. (1994). How effective is drug abuse resistance education? A meta-analysis of Project DARE outcome evaluations. *American Journal of Public Health, 84*, 1394-1401. doi: 10.2105/AJPH.84.9.1394

Esbensen, F.-A., Osgood, D. W., Taylor, T. J., Peterson, D., & Freng, A. (2001). How great is G.R.E.A.T.? Results from a longitudinal quasi-experimental design. *Criminology and Public Policy, 1,* 87-118. doi: 10.1111/j.1745-9133.2001.tb00078.x

Gottesman, D., & Schwarz, S. W. (2011). *Juvenile justice in the U.S.: Facts for policymakers.* New York, NY: Columbia University, Mailman School of Public Health, National Center for Children in Poverty. Retrieved from http://www.nccp.org/publications/pub_1038.html

Henry, S. (2000). What is school violence? An integrated definition. *Annals of the American Academy of Political and Social Science, 567*, 16–29. doi: 10.1177/000271620056700102

Jurney, C. (2013). Elementary school resource officers now permanent. *The Daily Tar Heel*, September 18. Last accessed online at: http://www.dailytarheel.com/article/2013/09/resource-officer-0917

Krug, E. G., Mercy, J. A., Dahlberg, L. L., & Ziwi, A. B. (2002). The world report on violence and health. *Lancet 360*(9339), 1083-1088. doi: 10.1016/S0140-6736(02)11133-0

Lipsey, M. (1992). Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects. In T. D. Cook, H. Cooper, D. S. Cordray, H. Hartmann, L. V. Hedges, R. J. Light, T. A. Louis, & F. Mosteller, (Eds.), *Meta-analysis for explanation: A casebook* (p. 83-127). New York, NY: Russell Sage Foundation.

Mark, M. M., & Reichardt, C. S. (2004). Quasi-experimental and correlational designs: Methods for the real world when random assignment isn't feasible. In C. Sansone, C. C. Morf, & A. T. Panter, (Eds.), *The SAGE handbook of methods for social psychology* (p. 265-286). Thousand Oaks, CA: Sage Publications, Inc.

Mytton, J., DiGuiseppi, C., Gough, D., Taylor, R., & Logan, S. (2006). School-based secondary prevention programmes for preventing violence. *Cochrane Database of Systematic Reviews, 3*, Article No. CD004606. doi: 10.1002/14651858.CD004606.pub2.

National Association of School Psychologists (2013). *Research on school security: The impact of security measures on students*. Bethesda, MD: Author. Retrieved from http://www.nasponline.org/advocacy/schoolsecurity.pdf

Neiman, S. (2011). *Crime, violence, discipline, and safety in U.S. public schools: Findings from the school survey on crime and safety: 2009–10 (NCES 2011-320)*. Washington, DC: U.S. Department of Education, National Center for Education Statistics, U.S. Government Printing Office.

Petrosino, A., Guckenburg, S., & Fronius, T. A. (2012). 'Policing schools' strategies: A review of the evaluation evidence. *Journal of MultiDisciplinary Evaluation, 8*(17), 80-101. Retrieved from http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/337

Petrosino A., Morgan, C., Fronius, T. A., Tanner-Smith, E. E., & Boruch, R. F. (2012). Interventions in developing nations for improving primary and secondary school enrollment of children: A systematic review. *Campbell Systematic Reviews, 19.* doi: 10.4073/csr.2012.19

Petrosino, A. (2005). Twenty years of D.A.R.E. and counting: A brisk history of evaluation

evidence and policy in America's most popular drug prevention program. *Japanese Journal of Sociological Criminology, 30*, 72-88.

Petteruti, A. (2011). *Education under arrest: The case against police in schools*. Washington, DC: Justice Policy Institute. Retrieved from http://www.justicepolicy.org/uploads/justicepolicy/documents/educationunderarrest_fullreport.pdf

Robers, S., Zhang, J., & Truman, J. (2012). *Indicators of school crime and safety: 2011* (NCES 2012-002/NCJ 236021). Washington, DC: National Center for Education Statistics, U.S. Department of Education, and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice.

Schultz, P. T. (2004). School subsidies for the poor: Evaluating the Mexican Progresa poverty program. *Journal of Development Economics, 74*(1), 199-250

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.

Shaw, M. (2004). *Police, schools, and crime prevention: A preliminary review of current practices*. Montreal, Canada: International Centre for the Prevention of Crime.

Smith, P. K. (2003). *Violence in schools: The response in Europe*. London: Routledge.

UNESCO (2007). *Expert meeting. Stopping violence in schools: What works?* Paris, France: UNESCO.

Weisburd, D., Lum, C., & Petrosino, A. (2001). Does research design affect study outcomes in criminal justice? *Annals of the American Academy of Political and Social Sciences 578*(1), 50–70. doi: 10.1177/000271620157800104

West, S. L. & O'Neal, K. K. (2004). Project D.A.R.E. outcome effectiveness revisited. *American Journal of Public Health, 94*(6), 1027-1029. doi: 10.2105/AJPH.94.6.1027

REVIEW AUTHORS

Lead review author:

Name:	**Anthony Petrosino**
Title:	Senior Research Associate
Affiliation:	WestEd
Address:	300 Unicorn Park Drive, 5th Floor
City, State, Province or County:	Woburn, MA
Postal Code:	01801
Country:	USA
Phone:	781-481-1117
Mobile:	
Email:	apetros@wested.org

Co-authors:

Name:	**Sarah Guckenburg**
Title:	
Affiliation:	WestEd
Address:	300 Unicorn Park Drive, 5th Floor
City, State, Province or County:	Woburn, MA
Postal Code:	01801
Country:	USA
Phone:	781-481-1116
Mobile:	
Email:	sgucken@wested.org

Name:	**Trevor Fronius**
Title:	
Affiliation:	WestEd
Address:	300 Unicorn Park Drive, 5th Floor
City, State, Province or County:	Woburn, MA
Postal Code:	01801
Country:	USA
Phone:	781-481-1134
Mobile:	

Email:	tfroniu@wested.org

ROLES AND RESPONSIBLIITIES

- Content: Anthony Petrosino, Sarah Guckenburg, and Trevor Fronius
- Systematic review methods: Anthony Petrosino
- Statistical analysis: Anthony Petrosino
- Information retrieval: Anthony Petrosino, Sarah Guckenburg and Trevor Fronius

SOURCES OF SUPPORT

Campbell Collaboration Education Group (via the University of Arkansas), $10,500.

DECLARATIONS OF INTEREST

The authors conducted a systematic search of the literature for a scoping study in 2011 for the U.K. National Police Improvement Agency, via George Mason University's Center for Evidence-based Criminology. The results were published in the *Journal of MultiDisciplinary Evaluation* (Petrosino, et al. 2012). Neither of those publications, however, made any definitive statement about the effectiveness of policing school strategies; instead, they focused on the quality of the evaluation designs.

PRELIMINARY TIMEFRAME

The review will be submitted in July 2014.

PLANS FOR UPDATING THE REVIEW

The review will be updated within three years after it is published.

AUTHOR DECLARATION

Authors' responsibilities

By completing this form, you accept responsibility for preparing, maintaining and updating the review in accordance with Campbell Collaboration policy. The Campbell Collaboration will provide as much support as possible to assist with the preparation of the review.

A draft review must be submitted to the relevant Coordinating Group within two years of protocol publication. If drafts are not submitted before the agreed deadlines, or if we are unable to contact you for an extended period, the relevant Coordinating Group has the right to de-register the title or transfer the title to alternative authors. The Coordinating Group also has the right to de-register or transfer the title if it does not meet the standards of the Coordinating Group and/or the Campbell Collaboration.

You accept responsibility for maintaining the review in light of new evidence, comments and criticisms, and other developments, and updating the review at least once every five years, or, if requested, transferring responsibility for maintaining the review to others as agreed with the Coordinating Group.

Publication in the Campbell Library

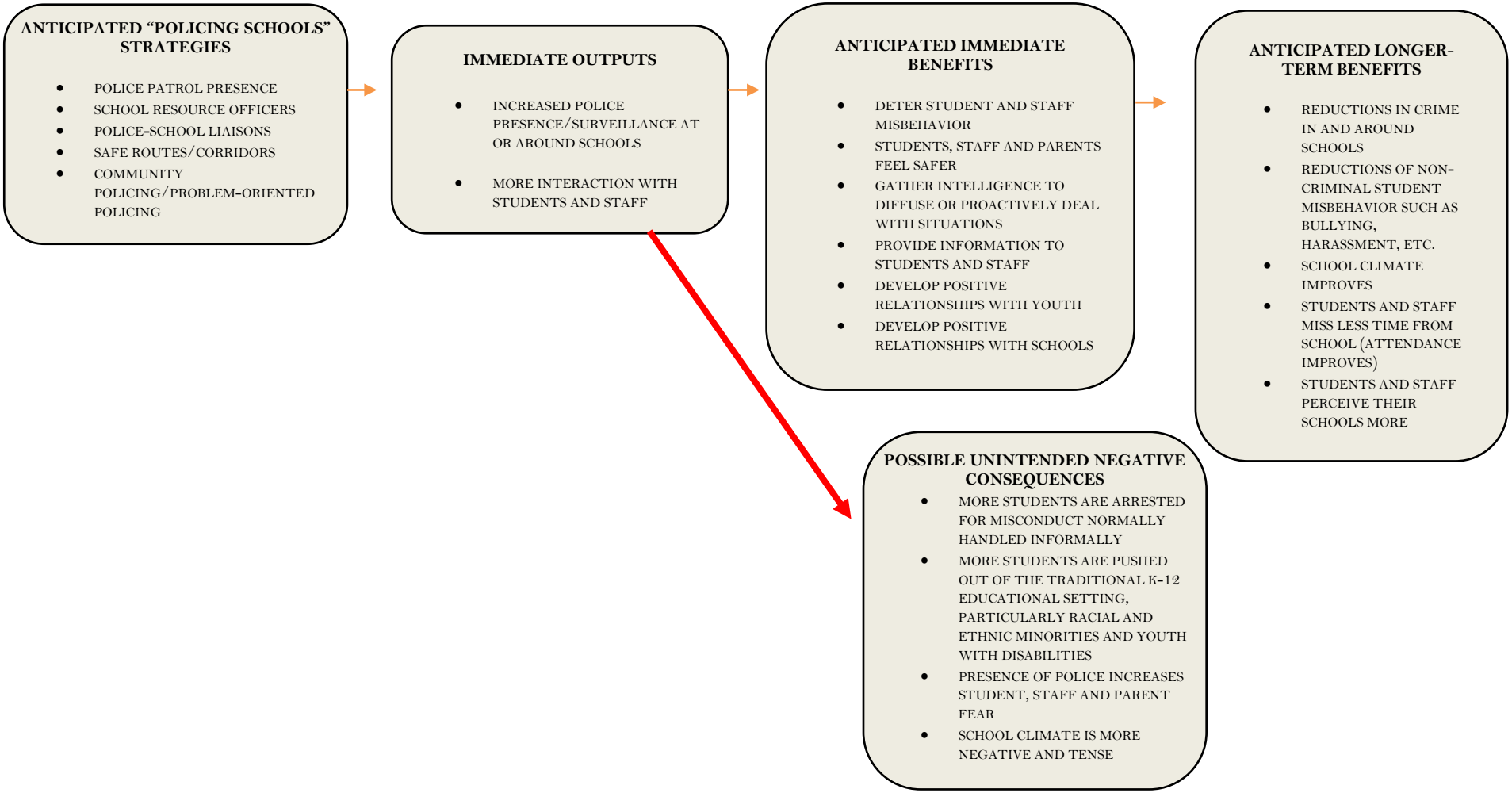
The support of the Coordinating Group in preparing your review is conditional upon your agreement to publish the protocol, finished review, and subsequent updates in the Campbell Library. The Campbell Collaboration places no restrictions on publication of the findings of a Campbell systematic review in a more abbreviated form as a journal article either before or after the publication of the monograph version in *Campbell Systematic Reviews*. Some journals, however, have restrictions that preclude publication of findings that have been, or will be, reported elsewhere and authors considering publication in such a journal should be aware of possible conflict with publication of the monograph version in *Campbell Systematic Reviews*. Publication in a journal after publication or in press status in *Campbell Systematic Reviews* should acknowledge the Campbell version and include a citation to it. Note that systematic reviews published in *Campbell Systematic Reviews* and co-registered with the Cochrane Collaboration may have additional requirements or restrictions for co-publication. Review authors accept responsibility for meeting any co-publication requirements.

I understand the commitment required to undertake a Campbell review, and agree to publish in the Campbell Library. Signed on behalf of the authors:

Form completed by:  **Date: 12/20/13**

APPENDICES

Appendix A. General theory of change for policing schools strategies



Appendix B. List of bibliographic databases to be searched

Academic Search Premiere
The Alcohol and Alcohol Problems Science Database (ETOH)
American Periodical Series Online
Bibliography of Nordic Criminology/Criminal Justice in Denmark
British Public Library Integrated Catalog
Canadian Evaluation Society Grey Literature
California Peace Officers Standards and Training Law Enforcement Archives
Chalk's E-Library
Chicano Database
Claseperiodica Abstracts (Caribbean and Latin America)
Cochrane Library: Cochrane Central Controlled Trial Register
Cochrane Library: Cochrane Health Technology Assessment Database
Cochrane Library: National Health Service Economic Evaluations
Conference Papers Index
Criminal Justice Abstracts
Database of African theses and Dissertations
Database of Research in International Education
EBSCO Master File
EBSCO Mega-file
EBSCO Military and Government Collection
EBSCO SOCINDEX
Econlit
ECONPAPERS
Education Administration Abstracts
Education Resources Information Center (ERIC)
Education Full-Text
Education Retro Index
Educators Reference Complete InfoTrac
Expanded Academic ASAP Plus
Family and Society Studies Abstracts
First Search OCLC
General OneFile Infotrac
Google
Google Scholar
Homeland Security Digital Library
Index to Current Urban Documents
Index to Foreign Legal Periodicals
International Bibliography of the Social Sciences
ISI Web of Knowledge/Social Science Citation Index
JRSA ISAR
JSTOR
Medline
National Bureau of Economic Research Working Papers
NCJRS
Ovid Full-Text Journals and Ovid Books
Policy Archive
Policy File
ProQuest Dissertations
Psychology and Behavioral Sciences Collection

Psychological Abstracts (PsycInfo)
Public Affairs Information Service (PAIS)
Public Affairs Information Service (PAIS) International
Race Relations Abstracts
Sage Criminology Full-Text
Selected Periodicals Index Online
Social Service Abstracts
Social Work Abstracts
Sociological Abstracts (Sociofile)
SSRN Electronic Library
Theses Canada
UK and Ireland Dissertations and Theses
Urban Studies Abstracts
World Bank Documents
Worldwide Political Abstracts

Appendix C. Coding Instrument

C2 Education Review: Policing Schools
CODING INSTRUMENT

Coder:

- ❑ Sarah Guckenburg
- ❑ Trevor Fronius
- ❑ Anthony Petrosino
- ❑ Alexis Stern
- ❑ Other _____

Citation for Primary Document:

_____
_____
_____
_____

I. RESEARCHER AND STUDY CHARACTERISTICS

What year was the document was published? _____

What was the type of primary document?

- o Book
- o Book Chapter
- o Government Report
- o Journal (peer reviewed)
- o Open-access electronic journal
- o Dissertation
- o Unpublished (tech report, conference paper)

How many documents were considered in coding this study? _____

What state or country did the study take place? _____

What was the setting for the study? _____

What other information was provided on the context for the evaluation? (This can include the rationale for the study; more about the setting; anything to help us learn more about the context)

_____
_____
_____

Indicate any inclusion/exclusion criteria for the units of analysis in the study

Who funded this study? (Should be in the acknowledgment section)

What were the relationships of the authors to the development and/or implementation of the treatment?

II. STUDY METHODS AND METHODOLOGICAL QUALITY

Was random assignment used to assign groups? (Yes/No)

At what level was randomization conducted? _____

How was the randomization specifically done? _____

Were any randomization problems (e.g., contamination, crossovers) noted? (Yes/No)

If yes, please detail those problems below:

How did investigators deal with randomization problems?

If random assignment was not used, what quasi-experimental method was used to equate groups? (e.g., matched comparison schools; post-hoc statistical matching of individuals; regression discontinuity; propensity scores; etc.)

Were any problems with non random assignment noted? (Yes/No)

If so, what were they? _____

How did investigators deal with non-random assignment problems?

Where did comparison group come from? _____

At what level was non-random assignment made? _____

Were any substantive differences in pretests of group equivalence noted? (Yes/No)

If yes, please detail those differences below:

_____
_____
_____
_____
_____

Was overall attrition problem from originally assigned sample noted? (Yes/No)

Was differential attrition noted? (Yes/No)

<u>If yes</u>, please detail those problems below (especially the magnitude of attrition, both from original sample and differentially between treatment and control groups):

_____
_____
_____
_____
_____
_____

How was attrition dealt with by investigators?

_____
_____
_____

III. INTERVENTION AND CONTROL CONDITIONS

Number of groups in the study: _____

Rationale for selecting intervention and control contrast if multiple groups:

_____
_____

List excluded study groups with brief description:

_____

_____
_____

Describe the intervention below, with particular attention to the "dosage" of the treatment:

_____
_____
_____
_____
_____

How many cases were randomized or assigned to this group? _____

Were there any program implementation problems described by investigators? (Yes/No)

Detail fidelity problems below:

_____
_____
_____
_____
_____

Please detail program theory (or mechanisms for why it should work. Also include any post-hoc information from study on why the program worked or did not work**):**

_____
_____
_____
_____
_____
_____

What is the control or comparison condition?

- o No Treatment Group
- o Wait-List Control
- o Treatment as Usual Group
- o Placebo
- o Lesser but Innovative Treatment

Describe the control or comparison condition (including "dosage" if applicable):
_____
_____
_____
_____
_____

How many cases were randomized or assigned to this group? _____

IV. PARTICIPANTS IN THE STUDY

Type of school _____

Age/school level/grade _____

Percentage of participants that were female _____

Percentage of participants that were white _____

Poverty/SES _____

Other data on participants: _____
_____
_____
_____

V. OUTCOMES

(SEE NEXT PAGE)

Include all data on treatment and control, including results, sample sizes used in analysis, the statistical technique, whether regression-adjusted or not, (and if so, what controls were used), statistical significance and probability level.

Outcome	**First Effect (Months:_____)**	**Last Effect (Months:_____)**
Crime/Behavior		
Perceptions		
Learning Outcomes		
Other		

Please detail all subgroup effects below:

Provide any information on qualitative data in the study:

Please detail all cost/economic information below:

ANY OTHER COMMENTS ON THE PROGRAM OR EVALUATION (use bullets)