The Effectiveness of Volunteer Tutoring Programs:
A Systematic Review for the Campbell Collaboration
Education Review Group

Revised Protocol • December 22, 2005

Reviewers:

Ginger R. Albin
Higher Education Ed.D Student
University of Arkansas
257 Graduate Education Building
Fayetteville, AR 72701
Phone: (479) 575-3586
galbin@uark.edu

Joshua H. Barnett
Public Policy Ph.D Student
University of Arkansas
15 West Avenue Annex
Fayetteville, AR 72701
Phone: (479) 575-4930
jhb05@uark.edu

Virginia L. Blankenship
Research Associate, Office for Education Policy
University of Arkansas
15 West Avenue Annex
Fayetteville, AR 72701
Phone: (479) 575-4930
vhudson@uark.edu

Dr. George Denny
Professor, Education Leadership, Counseling and Foundations
University of Arkansas
231 Graduate Education Building
Fayetteville, AR 72701
Phone: (479) 575-7320
gdenny@uark.edu

Dr. Gary W. Ritter – Lead Reviewer
Assistant Professor, Education and Public Policy
University of Arkansas
241 Graduate Education Building
Fayetteville, AR 72701
Phone: (479) 575-4971
Fax: (479) 575-2492
I. Background for the Review

In many cultures, the oldest form of teaching was provided to children by in-home tutors or private instructors (Shanahan, 1998). Tutoring remains a popular form of instruction worldwide, and the effectiveness of tutoring as a pedagogical method has been documented extensively in various strands of the educational literature (see, for example, Cohen, Kulik, & Kulik, 1982, and Wasik & Slavin, 1993).

During the 1970s, U.S. schools began relying more on peer tutoring (also known as student-to-student or cross-age tutoring) as a way to efficiently use scarce resources in a period of teacher shortages (Rekrut, 1994). The next two decades also witnessed an increased interest in tutoring programs staffed by adult volunteers for a variety of reasons: (1) increased public concern with the quality of education after the U.S. Department of Education’s release of “A Nation at Risk” in 1983; (2) rising interest in community service in the 1990s, and (3) the encouraging results from effective yet costly programs that employ professional tutors. By 1987, the National Research Council estimated that there were over one million volunteer tutors who donated an average of four hours per week in the nation’s public schools. The survey found that three-fourths of public elementary schools in the United States reported the involvement of volunteers, with each school having an average of 24 volunteers (Michael, 1990).

This heightened interest in volunteerism may be most evident at institutions of higher education around the nation, many of which have entered into tutoring partnerships with nearby school districts, particularly those in the inner-city. During the 1987-1988 school year, the U.S. Department of Education sponsored a survey to identify college-sponsored programs that used college students as tutors or mentors for preschool, elementary, or secondary students. The survey found that nearly one-third (29%) of all colleges or universities hosted a program of this type, with over 70,000 college students serving 240,000 K-12 students in over 1,000 schools (Cahalan & Farris, 1990). In two-thirds of these programs, tutoring was the primary focus.

More recently, the growth of these university partnerships has accelerated due to the America Reads Challenge, a nationwide tutoring initiative launched in 1997 by President Clinton. At that time, nearly 800 universities and colleges throughout the nation had already pledged to commit work-study slots for college students to serve as tutors for elementary school children (White House, 1997). By 1999, nearly 1,200 colleges and universities committed to placing work-study students as tutors in public schools. President Clinton’s 1999 budget proposal also included $140 million to establish programs matching university-based mentors with students in schools that had very high dropout rates and high concentrations of poor students. Tutoring during the school day was among the many services the mentors would provide (Seelye, 1998). Then-U.S. Education Secretary Richard Riley described community and school partnerships as “keys to
educational excellence and the only way to achieve sustained involvement over time” (Sandham, 1998).

As a result of the America Reads Challenge, state leaders have become increasingly interested in providing tutoring programs for elementary school children, and numerous local tutoring initiatives are now receiving increased support (U.S. Department of Education 1996; 1997). For example, in the late 1990s, Governor Gary Locke of Washington called on state legislators to create a corps of 25,000 volunteers to help tutor students in grades two through five (Trotter, 1998). Since 1998, this program has enabled 11,000-trained volunteers to give intensive tutoring to 22,000 elementary students statewide and has been shown to have a positive impact on students’ reading ability (Governor Locke’s Major Accomplishments 1997-2002, n.d.). Other states, such as Arkansas, also allocate $500,000 annually for nonprofit volunteer tutoring efforts (Southern Regional Education Board, 2000).

Research on Tutoring Programs

Despite the increased interest in and support for tutoring programs over the past few decades, the expansion of programs that use non-professional, adult volunteer tutors has yet to be matched by a supporting research base. In the 1970s and early 1980s, most tutoring research in the United States focused on the impacts of peer or cross-age tutoring (Shanahan, 1998). In 1982, Cohen, Kulik, and Kulik published a well-known and often-cited review of this early body of tutoring research in the American Educational Research Journal. The authors noted that hundreds of reports on tutoring had been written by teachers and researchers, some based on scientifically-sound, experimental design evaluations, and others more informal and subjective. According to the authors, the four major reviews of the research up to 1982 used “relatively informal narrative and box score techniques” (p. 234), and each concluded that peer tutoring can help improve the academic success of young students. ¹

The reviewers argued that more formal review methods were necessary and used these early reviews of the literature, along with other studies, as the basis for a meta-analysis of findings from 65 independent evaluations of tutoring programs in schools. The vast majority of the programs reviewed employed students—either peer or older students—as tutors. Their meta-analysis concluded that the message from the educational literature was clear: “These programs have definite and positive effects on the academic performance and attitudes of those who receive tutoring” (Cohen et al., 1982, p. 244).

During the 1980s and 1990s, researchers began focusing more on specialized interventions aimed at improving the academic achievement of the lowest-achieving children, most notably Reading Recovery and Success for All (Shanahan, 1998; Wasik & Slavin, 1993). Both programs include one-on-one tutoring by professional tutors and are perceived as effective by many in the research community (Wasik, 1998). However, because the expense of employing professional tutors limits the number of children who can be served by these interventions, several programs have been created in recent years that use adult volunteers or para-professional as tutors.

¹ The four reviews are: Devin-Sheehan, Feldman, & Allen, 1976; Ellson, 1976; Fitz-Gibbon, 1977; and Rosenshine & Furst, 1969.
In our initial review of the existing literature, we have found that there are several obstacles that limit the number of rigorous evaluations of the effects of volunteer tutoring programs: (1) good evaluations are expensive and require highly-qualified methodologists; however, volunteer programs generally operate on small budgets and prefer to allocate all available resources to program delivery rather than to evaluation; (2) program operators may view volunteer resources as “free” and see no need for rigorous evaluations; (3) program operators may be resistant to any evaluation method, such as random assignment, that might alter program practice or raise questions regarding who is receiving the services; and (4) program operators may be cautious of external evaluation due to a belief that student achievement, particularly as measured by test scores over a short time period, is not the best indicator of program effectiveness.

As a result of these research limitations, not only is there a scarcity of impact evaluations of volunteer tutoring programs, but there are also many weaknesses in the evaluations that have been conducted. Recent studies and reviews have highlighted the strong selection bias often found in quasi-experimental studies, particularly of programs with voluntary participation of students (Glazerman, Levy, & Myers, 2003; Wilson & Lipsey, 2001). In addition, the reporting of methodology and findings in many studies often leaves out important variables and procedures which may have a large impact on the effectiveness of the programs. Therefore, the nature of volunteer tutoring programs presents many challenges for researchers and reviewers of empirical studies.

Recent Reviews of Evidence

In our early searches, we have identified four major recent reviews of research on the impact of volunteer tutoring programs on student outcomes: Topping & Hill (1995), Wasik (1998), Shanahan (1998), and Elbaum et al. (2000).

In 1995, Topping and Hill provide interesting background on the research related to volunteer tutoring in a chapter contributed to the book, Students as tutors and mentors (Goodlad, 1995). The chapter presents a review of the evaluation research around the world related to the effectiveness of college students as tutors for schoolchildren.

Wasik (1998) reviewed studies of 17 programs that used volunteer tutors to help improve students’ reading abilities. While the evidence suggested that volunteers can indeed help many children improve their reading skills, the results were quite varied across programs. Furthermore, only two of the 17 programs reviewed compared students’ achievement with that of a control group, which makes the results less reliable. Wasik’s findings are consistent with those of the National Research Council’s 1998 report, Preventing Reading Difficulties in Young Children, which concluded that there is “[n]o evidence confirming that [volunteers] are able to deal effectively with children who have serious reading problems” (p. 238). Nevertheless, we suspect that there are many other evaluations not included in either of these non-systematic reviews, which could have altered the findings.

In contrast, Shanahan’s (1998) review of the research on volunteer tutoring found that, despite many limitations, these programs can indeed be effective in improving student achievement. However, Shanahan offers little detail about the methodology used in his review, and, as with the reviews listed above, the results of several important studies are omitted.
Most recently, Elbaum et al. (2000) reported on a meta-analysis in the Journal of Educational Psychology focused on the effectiveness of one-to-one tutoring programs for improving reading ability in elementary students at risk for reading failure. The authors reviewed 29 studies involving 42 samples of students between 1975 and 1998 and found that trained tutors can help students improve in reading skills.

The limited and often-conflicting evidence on the effectiveness of volunteer tutoring programs leaves us with many questions. For example, how do differences in participants (i.e., socioeconomic status, gender) impact the effectiveness of the programs? And how do differences in tutoring programs (i.e., extent of tutor training, length of program) affect the results?

II. Objectives of the Review

Because relatively little attention has been paid to evaluating the effectiveness of volunteer tutoring programs, little is known about how to best utilize the many volunteers working in our schools. However, in recent years, some researchers have begun to address these concerns by designing and conducting rigorous evaluations of these programs. The objective of this systematic review is to gather, summarize, and integrate the empirical research on volunteer tutoring programs in order to help policymakers, educators, parents, and other stakeholders understand the evidence on which programs work best for which students and how they can best be implemented. In particular:

1) Do volunteer tutoring programs impact academic outcomes (either reading or math) of students? If so, which programs, and to what extent?

2) For which subgroups of students are volunteer tutoring programs most or least beneficial?

3) What are the distinguishing characteristics of the most successful volunteer tutoring programs?

The need for this review has become even more pressing since the implementation of the federal No Child Left Behind (NCLB) Law of 2001, which requires schools to use evidence-based research in designing and implementing educational programs and policies. The accountability portion of the law also requires schools receiving federal Title I funding that fail to make adequate yearly progress (AYP) for two or more consecutive years to offer and pay for supplemental education services (such as tutoring) for their students. Many of these supplemental service providers employ volunteers as tutors, either in part or in total. Millions of public school students across the country are now eligible to receive these services (Center on Education Policy, 2005).

However, few states or districts are prepared to select, monitor, and evaluate tutoring programs and volunteers (Center on Education Policy, 2005; Gewertz, 2005). Therefore, this review will not only fill in an important gap in the research base, but more immediately and practically,
could also help guide policymakers and educators implement and evaluate effective programs and policies across the country.

III. Methods

Our pilot review of the research on volunteer tutoring programs began in 2001 thanks to the Campbell Collaboration’s “test-bed” effort. The lead reviewer has since undergone additional training in conducting C2 systematic reviews, and a draft review protocol has already been submitted to C2 and received several reviews. The reviewers’ concerns primarily focused on providing details about our strategies for data analysis. For the most part, the reviewers deemed our search strategies to be appropriate. Thus, we proceeded with the preliminary search of relevant studies, which resulted in roughly 12 randomized controlled trials (RCTs) that met our inclusion criteria. We now intend to update the search to find new evaluations or others missed in the prior search process, then analyze and synthesize all of the included studies. Below is a description of our review methodology.

a. Criteria for Inclusion and Exclusion of Studies

Only empirical studies that meet the following criteria will be kept:

Types of studies

Only randomized field trials will be included in the review. Furthermore, we will exclude designs in which more intervention is compared to less intervention. That is, we will only include studies where a control group, which has an absence of any intervention, is used. Quasi-experimental studies that employ treatment and control groups matched on pre-tests of key outcome variables will not be included in this review, but will be clearly identified so that we can easily include these in the review at a later date if appropriate. Pretest-posttest studies, or those in which a treatment group is compared to another treatment group, will not be included.

Studies: must be published later than 1985. Only English-language studies of programs conducted in the United States will be considered, due to the limited resources for this review. Furthermore, we will exclude studies of programs that are especially designed to address the needs of students with limited English proficiency (LEP), since such specialized programs are not representative of most volunteer tutoring programs for elementary and middle school students.

Types of participants and interventions

Only studies of programs involving adult, non-professional tutors will be included. Although these tutors are almost always referred to as “volunteers” in the literature, those programs that pay a small stipend to tutors (such as undergraduate tutors who are tutoring as part of a work-study program) will also be included. However, tutors who are salaried district employees but are tutoring after school will be excluded. In terms of the tutees, only studies of programs that serve students in grades K-8 (elementary and middle school) will be considered, since this is the population typically served by volunteer tutoring programs, and such programs are fundamentally different than those provided to high school students.
The intervention must feature regular face-to-face tutoring sessions with an academic focus for at least one month in duration. The face-to-face requirement means that we will not include studies where the intervention was conducted online or via telephone. However, programs may also include other components in addition to an academic focus (such as a summer program that offers art or swimming lessons in addition to tutoring); however, the program must have at least one academic outcome measure.

**Types of outcome measures**

Most volunteer tutoring programs have been evaluated based on various outcome measures (i.e., Iowa Test of basic Skills, Basal Word Recognition, academic grades). The standardized exam content areas of interest include reading, writing, and mathematics. The review will consider all outcome measures related to student achievement, including distal outcomes (ones we actually want to influence, such as school grades or standardized achievement measures), as well as proximal outcomes (intermediate measures that might be influenced by tutoring and then might lead to improved outcomes in the future, such as student attendance rates). The reviewers will then calculate effect sizes for each measure in an attempt to make the data as comparable as possible.

**Assessment of study quality**

The quality of each study (and its reporting) will be assessed according to several characteristics, including: 1) the transparency of the study; that is, the clarity with which the investigators report the random assignment procedures; 2) the integrity of the random assignment design and whether investigators address violations of the design; and 3) the existence of high levels of attrition (particularly, differential attrition between treatment and control groups) of either tutees or tutors from the sample initially randomized. Other studies with egregious methodological problems will also be eliminated from inclusion in the final review.

In addition to study quality, we will also check to see how the results are reported. The results must either include the effect size or include enough information for us to be able to calculate the effect size. In cases where the reported data seem inconsistent with other information reported, we will exclude the study. For example, in one study retained for full coding, an ANOVA was conducted on two groups with a total sample of 55 participants, but the reported degrees of freedom were 1,92. In such cases, where the reported data seems inconsistent, we will exclude the study.

**b. Search Strategy for Identifying Relevant Studies**

For this systematic review, titles of studies on volunteer tutoring programs will be identified using several methods. First, the reviewers will search the EbscoHost Research Database using the following databases: Academic Search Premier; Primary Search; Professional Development Collection; Middle Search Plus; Psychology and Behavioral Sciences Collection; PsycINFO; Sociological Collection, ERIC (Education Resources Information Center), and Proquest Digital Dissertations.

The initial search will include the following key words in various combinations and truncations: “Volunteer or mentor or tutor* or tutorial programs,” and “elementary or primary education or
middle school students or junior high school students or early intervention,” and “control or random or experiment or evaluation or program not peer.” We will also hand-search the table of contents of several major journals (for years 2003-2005) that are most relevant to our study, including: *Education Next*, *Education Policy Analysis Archives* (EPAA), *Educational Evaluation and Policy Analysis* (EEPA), *Reading Research Quarterly* (RRQ), and *Review of Educational Research* (RER). Due to the limited resources available for this review, only English-language publications and databases will be consulted. Finally, the resulting list of articles will be augmented by other research studies referenced in four widely-cited reviews on volunteer tutoring listed above: Elbaum et al., 2000; Wasik, 1998; Shanahan, 1998; and Topping & Hill, 1995.

The reviewers will also consult with several sources in order to refine the search process, including an information specialist or librarian, a reading specialist, and the Campbell Collaboration’s Information Retrieval Policy Brief (Rothstein, Turner, & Lavenberg, 2004). Studies will be retrieved primarily from the University of Arkansas Library System, Interlibrary Loan, University Microfilms, and the databases listed above. All study titles and inclusion decisions will be will be documented and managed using Excel software in order to maintain accuracy and consistency among the reviews. When possible, PDF files of all articles will be saved in a central network folder; hard copies of these and print-only articles will also be kept on file.

The list of study titles generated in this process will then be narrowed through a review of the studies’ abstracts by at least two reviewers. Once the abstracts have been retrieved and reviewed, the full text of all studies chosen by both reviewers in this process will then be read and coded for further analysis.

c. Description of Methods Used in the Component Studies

Typical evaluations of volunteer tutoring programs employ either experimental or quasi-experimental designs and include testing of treatment and control groups both before and after the intervention. The most common outcome investigated is reading achievement, usually measured by standardized test scores.

One study which exemplifies the methods most commonly used in tutoring evaluations examines a one-on-one tutoring program called “Sound Partners” for inner-city minority students (Vadasy, Jenkins, Antil, Wayne, & O’Connor (1997). The authors employed a stratified random assignment procedure based on students’ reading pretest scores. Compared to most studies, the authors present greater detail on the program and student characteristics and methodology used, and even include effect sizes for each outcome measure. They found statistically significant positive effects in spelling and word segmentation measures for tutored students compared to non-tutored students, but found no differences between the treatment and control group in reading achievement. It should be noted that the evaluators in this study were also the program developers, and like many studies, the evaluation includes a small sample size.

d. Criteria for Determination of Independent Findings
As mentioned above, many studies report results on multiple outcome measures (i.e., reading scores, math scores, attendance) and often include several variations of the same measure (i.e., teacher, parent, and/or student reports of math improvement). Effect sizes for each measure will be extracted and coded in our analysis. The methods for maintaining statistical independence during analysis in cases where multiple effect sizes are available are described below.

e. Details of Study Coding Strategies

At least two reviewers will independently extract and code data from each full article in a coding guide which will include the following:

- Study citation & author affiliations
- Study sponsor & relation to program
- Peer-reviewed or non-peer-reviewed
- Program objective & rationale
- Program location & setting (urban vs. rural, in school vs. after school, etc.)
- Time frame of program & study
- Sampling & assignment procedure (i.e., random? quasi-random?)
- Description of tutors, training & compensation (if applicable)
- Student characteristics in both experimental and control groups (grade level, gender, ethnicity, & socioeconomic status)
- Program characteristics (i.e., length of tutoring sessions, highly structured vs. unstructured sessions, resources provided)
- Attrition of tutors and tutees
- Analytical techniques (i.e., multiple regression analysis?)
- Outcome measures used as indicators of student achievement for tutees (and subgroups, if applicable)
- Results of the study
- Methodological weaknesses and criticisms of study designs
- Inclusion decision for systematic review (met criteria or not)

Inter-rater agreement, or coding reliability, will be assessed for all studies, and resolution of coding disagreements will be resolved by meeting and discussing contested items. Only data with perfect agreement will be entered for each study. A detailed codebook can be found in the Appendix of the protocol. We will also consider the coding schemes used in recent meta-analyses (i.e., Elbaum, Vaughn, Hughes, & Moody, 2000).

f. Statistical Procedures and Conventions

Once we have identified and compared all outcomes measured in the included studies, we will select the appropriate effect size metric for the meta-analysis. We will most likely use the standardized mean difference effect size statistic (the difference between the treatment and control group means on an outcome variable divided by their pooled standard deviations) for each outcome measure (Lipsey & Wilson, 2001). When means and standard deviations are not
reported, we will attempt to estimate the effect sizes using the procedures recommended by

In addition to effect size values, details on the methods and procedures (i.e., design, attrition), the
intervention (i.e., duration, setting), and the subject samples (i.e., age, gender, ethnicity) will be
coded for each study to determine the relationship between dependent and independent variables.

**Missing Data**

In the case of critical data not being reported in the studies, we will attempt to contact study
authors for the information. If we are unable to find the critical information, we will exclude
such studies from our analysis.

### Multiple Outcomes for Single Studies (and heterogeneous outcome measures)

It is likely (based on prior reviews) that the evaluation studies of volunteer tutoring programs
employ a variety of different outcome measures to assess program effectiveness. Some
evaluations use general reading scores on standardized exams while others use authentic
measures (e.g. how many words read correctly during oral reading session) while others use
state-level standardized assessments. Examples of standardized exams used include the Wide
Range Achievement Test (WRAT), the Stanford Achievement Test (SAT-9), and the Iowa Test
of Basic Skills (ITBS).

Because the different studies employ different outcomes in different ways, it may not be prudent
to calculate the “effect” of each individual study or the “overall effect” of all available studies.
Alternatively, so as to determine whether this type of intervention has a greater effect in any one
area, we will conduct separate meta-analyses of key outcome areas, such as standardized overall
reading, standardized reading sub-skills, or authentic measures of reading ability. If a study
measures a key outcome in several ways, we will ensure that each study only contributes one
data point to the analysis for each key outcome in order to ensure that no individual study is
unduly “weighted” in the meta-analysis.

### Software and Resources

All study coding and data management will be done using Excel software. In order to conduct
the meta-analysis, the authors will use Comprehensive Meta-Analysis software developed by
Biostat. Other researchers with expertise in meta-analysis will be consulted throughout this
process, as needed.

### Sensitivity Analysis

The reviewers will also conduct a sensitivity analysis that compares the different results of
subgroups of studies, including: 1) studies reporting on direct vs. indirect outcome measures; 2)
studies of highly structured vs. unstructured programs; and 3) studies of programs focusing on
reading vs. those focused on other academic subjects. As previously noted, we will then calculate
effect sizes in attempt to compare the data between studies as best as possible.

### Treatment of Qualitative Research
This review focuses on studies involving randomized controlled trials because of the advantages this analysis has over other methods of analysis; therefore, no qualitative studies will be coded for the purposes of this project.

**IV. Timeframe**

The reviewers intend to complete this systematic review by the end of 2005. As previously mentioned, the review team has already conducted a pilot review of the research on volunteer tutoring programs. Therefore, much of the research that will be used in the final review has already been collected and summarized. We will again review these studies for the accuracy of data extraction and analysis and add additional articles via the search strategies and inclusion criteria listed above. After all final articles are selected to be included in the review we will conduct a meta-analysis of the results extracted from each study.

The proposed project timeline is as follows:

<table>
<thead>
<tr>
<th>Event</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revised protocol</td>
<td>June 30, 2005</td>
</tr>
<tr>
<td>Updated search</td>
<td>August 15, 2005</td>
</tr>
<tr>
<td>Data extraction and analysis</td>
<td>October 15, 2005</td>
</tr>
<tr>
<td>Final report to Campbell Collaboration</td>
<td>December 31, 2005</td>
</tr>
<tr>
<td>Revisions to final report</td>
<td>January 31, 2006</td>
</tr>
<tr>
<td>Presentation at C2 conference in Los Angeles</td>
<td>February 2006</td>
</tr>
<tr>
<td>Publication of results</td>
<td>2006</td>
</tr>
</tbody>
</table>

**V. Plans for Updating the Review**

The authors will attempt to update the review approximately every five years, pending funding.

**VI. Acknowledgements**

The initial investigation into the field of volunteer tutoring was supported under the “Evaluation of the West Philadelphia Tutoring Project” conducted by faculty and staff at the University of Pennsylvania from 1998-2000. Funding for this project was provided by the Smith Richardson Foundation under Grant #9801-155 to the University of Pennsylvania and the University of Missouri. The draft protocol for the pilot review was completed with the help of many colleagues, including participants in the University of Pennsylvania’s Graduate School of Education seminar on conducting systematic reviews and meta-analysis, led by Dr. Rebecca Maynard.

The systematic review is now being conducted at and funded by the University of Arkansas. The results of this study do not necessarily represent the official opinion or policy of the Smith Richardson Foundation, the University of Pennsylvania, or the University of Arkansas.

**VII. Statement Concerning Conflict of Interest**
Under the advisement of Rebecca Maynard, Ritter (2000) previously completed an impact evaluation of a university-based tutoring program using random assignment that will be reviewed for its potential inclusion in the systematic review of the literature. Dr. Ritter will recuse himself from the decision to include this study in the meta-analysis.

Virginia Hudson has previously served briefly as both a paid-employee and volunteer tutor in various schools and community organizations, including the University of Arkansas’ Upward Bound program. She recently received tutor training at the Ozark Literacy Council but is not currently serving as a tutor.

The reviewers maintain that their prior limited experience with volunteer tutoring programs will not affect their view of research or final results. Furthermore, any potential bias related to our earlier work in this area should be counter-balanced by the explicit and transparent methods used in the systematic review.

VIII. References


IX. APPENDIX

DATA ABSTRACTION TEMPLATE

A. Context

A1. Location of program  rural  urban  suburban

A2. Setting
   Public school
   Religious institution
   Day care center
   Community center
   Other:

A3. Number of sites:

A4. Evaluation date:

A5. Organization responsible for program operation
   Local School
   School District
   Non-profit organization
   University
   Other:

A6. Program operation
   Days per week
   Hours per day
   Time of day
   Days per year

A7. Staff
   Total number
   Number per site
   Number who are classroom teachers
   Average number of teachers daily

A8. Participants
   Grades served
   Ages served
   Total number served
   Number per site
   Average daily attendance

A9. Program eligibility rules
   Subjective? Specify:
   Objective?
   Target population?
A10. Alternative programs for non-participants
   Yes
   No

A11. Program goals
   Academic
   Recreational
   Developmental
   Combination (list)

A12. List goals as program defines them:

A13. Description of the program activities, including the average treatment duration and variance.

A14. Connections made to school day?
   Yes
   No
   How?

A15. When do program activities occur (during the school day or weekend)?
   Describe:
   Describe:

A16. Other details about the context which were provided in the study and have not yet been captured.

B. Evaluation

B1. Authors and affiliations:

B2. Time period for evaluation:

B3. Evaluator’s relationship with intervention
   Program developer
   Program operator
   Advocates in favor of intervention
   Advocates opposed to intervention
   Other or not stated. Describe:
   External?

B4. Publication source:

B5. How was study identified?

B6. Peer reviewed journal?

B7. Implementation of Experimental Design (Level of Random Assignment)
   Constraints for random assignment? Y/N _____________ Describe
   Alternative treatment for control group? Y/N _____________ Describe
   Was there no-change control group? Y/N _____________ Describe
Fidelity of implementation:

B8. Sample Attrition
Attrition rate for treatment
Attrition rate for control
How do they deal with attrition analytically?

B9. Outcome measures
Were some measures given to treatment and control?
Describe:
Names of Standardized Outcome Measures
Names of Non-standardized Outcome Measures
Any evidence of validity or reliability?

C. Sample

C1. Total Sample Size

<table>
<thead>
<tr>
<th></th>
<th>Treatment</th>
<th>Control</th>
<th>Total Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age/Grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>Female</td>
<td></td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td>African American</td>
<td>Latino</td>
<td>White</td>
</tr>
<tr>
<td>Free/Reduced Lunch</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other demographic factor</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

D. Data analysis and results

D1. Describe statistical methods, including independent variables.

D2. Outcome measures (specify time period covered, if relevant)
Higher scores measure?
Table for Each Subgroup and Entire Sample

<table>
<thead>
<tr>
<th>Measure 1:</th>
<th>Measure 2:</th>
<th>Measure 3:</th>
<th>Measure 4:</th>
<th>Measure 5:</th>
<th>Measure 6:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Participant Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sdv Participant Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effect Size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Significance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

D3: Control variables?

D4: Please describe any issues or problems with the data analysis:

D5: Researcher’s Discussion of Results