Protocol:

**Dropout Prevention and Intervention Programs: Effects on School Completion and Dropout Among School-aged Children and Youth**

Sandra Jo Wilson, Mark W. Lipsey, Emily E. Tanner-Smith, Chiungjung Huang, & Katarzyna Steinka-Fry

Protocol submitted: April 16, 2010
Revised protocol submitted: August 23, 2010

**BACKGROUND**

With the expansion of regional and national economies into a global marketplace, education has even greater importance as a primary factor in allowing young adults to enter the workforce and advance economically, as well as to share in the social, health, and other benefits associated with education and productive careers. Dropping out of school before completing the normal course of secondary education greatly undermines these opportunities and is associated with adverse personal and social consequences. Dropout rates in the United States vary by calculation method, state, ethnic background, and socioeconomic status (Cataldi, Laird, & KewelRamani, 2009). Across all states, the percentage of freshman who did not graduate from high school in four years ranges from 13.1% to 44.2% and averages 26.8%. The status dropout rate, which estimates the percentage of individuals in a certain age range who are not in high school and have not earned a diploma or credential, is slightly lower. In October 2007, the proportion of noninstitutionalized 18-24 year olds not in school without a diploma or certificate was 8.7%. Males are more likely to be dropouts than females (9.8% vs. 7.7%). Status dropout rates are much higher for racial/ethnic minorities (21.4% for Hispanics and 8.4% for Blacks vs. 5.3% for Whites). Event dropout rates illustrate single year dropout rates for high school students and show that students from low-income households drop out of high school more frequently than those from more advantaged backgrounds (8.8% for low-income vs. 3.5% for middle income and 0.9% for high income students). The National Dropout Prevention Center/Network reports that school dropouts in the United States earn an average of $9,245 a year less than those who complete high school, have unemployment rates almost 13 percentage points higher than high school graduates, are disproportionately represented in prison populations, are more likely to become teen parents, and more frequently live in poverty (2009). The consequences of school dropout are even worse for minority youth, further exacerbating the economic and structural disadvantage they often experience.

School dropout has implications not only for the lives and opportunities of those who experience it, but also has enormous economic and social implications for society at large. For instance, the
National Dropout Prevention Center/Network (2009) reports that each annual cohort of dropouts costs the United States over $200 billion during their lifetime due to lost earnings and unrealized tax revenue; and even a 1% increase in high school graduation rates could save over $1 billion in incarceration costs. The Organisation for Economic Co-operation and Development (2009) has similarly documented the tremendous social and economic gains associated with secondary school completion in OECD member countries.

A relatively large number of intervention and prevention programs in the research literature give some attention to reducing dropout rates as a possible outcome. The National Dropout Prevention Center/Network, for instance, lists 192 “model programs.” Relatively few of those programs, however, bill themselves as dropout programs; many focus on academic performance, risk factors for dropout such as absences or truancy, or indirect outcomes like student engagement, but may also include dropout reduction as a program objective. The corresponding research domain includes evaluations of virtually any program provided to students for which dropout rates are measured as an outcome variable, regardless of whether they are billed as dropout programs. To represent the full scope of relevant research on this topic, all such programs should be considered in a review of dropout programs. However, because we are interested in summarizing the research on dropout programs that could be implemented by schools, we narrow our focus to programs that can be implemented in school settings or under school auspices.

There have been a handful of systematic reviews on the effects of prevention and intervention programs on school dropout and completion outcomes. However, the restrictive inclusion criteria and methodological weaknesses of these reviews preclude any confident conclusions about the effectiveness of the broad range of programs with dropout outcomes, or the potential variation of effectiveness for different program types or subject populations. For instance, the U.S. Department of Education’s What Works Clearinghouse report on dropout prevention found only 15 qualifying studies that reported outcomes on direct measures of staying in school or completing school (http://ies.ed.gov/ncee/wwc/reports/dropout/topic/#top). This report, however, restricted discussion to interventions in the United States and did not include a meta-analysis of program effectiveness or examine potential moderators of program effectiveness. Another review on best practices in dropout prevention summarized the results of 58 studies of dropout programs (ICF, International, 2008). That report presented effect sizes primarily for individual program types and did not examine potential moderators or examine the influence of study method on effect size. The report also presented a narrative review of important variables associated with implementation quality, but implementation quality was not analyzed in a meta-analysis framework.

Two other systematic reviews have focused on the effectiveness of prevention and intervention programs to reduce school dropout or increase school completion (Klima, Miller, & Nunlist, 2009; Lehr et al., 2003). In their review, Lehr et al. (2003) identified 17 experimental or quasi-experimental studies with enrollment status outcomes. This review was completed seven years ago, and thus does not include the most recent studies. The authors did not perform a meta-analysis because they felt that the dependent variables differed too greatly across studies to create meaningful aggregates. This circumstance prevented the authors from examining the differential effectiveness of programs with different treatment or participant characteristics, something we plan to do in the proposed systematic review. In a more recent review, Klima et al. (2009) identified 22 experimental or quasi-experimental studies with dropout, achievement, and truancy outcomes. However, this review excluded programs for general “at-risk” populations of students (e.g., minority or low socioeconomic status samples), as well as programs with general character-building, social-emotional learning, or delinquency/behavioral
improvement components. These exclusion criteria therefore limited the conclusions that could be drawn about the broader range of programs that aim to influence school dropout and completion outcomes. Further, this review only presented mean effect sizes for different types of interventions, and did not examine the potential variation of effects for different subject populations.

The findings of the Klima et al., (2009) and Lehr et al., (2003) reviews have some similarities. Both teams highlight the dearth of high-quality research on dropout programs, and mention especially the lack of key outcomes such as enrollment (or presence) at school and dropout. Both reviews demonstrate that some of the included programs had positive effects on the students involved. Lehr and her colleagues do not identify specific programs that were particularly effective or ineffective, but focus rather on implementation integrity as a key variable and emphasize the importance of strong methodologies for future research on dropout programs. Klima and colleagues conclude that the programs they reviewed had overall positive effects on dropout, achievement, and attendance/enrollment. They highlight alternative educational programs, such as schools-within-schools, as particularly effective. The Klima review also suggests that alternative school programs, that is, programs in separate school facilities, were ineffective. Overall, these two reviews identify several important potential moderators that will be included in the coding scheme for the proposed review. These include implementation quality, treatment modality, and whether programs are housed in typical school facilities or in alternative school locations.

**OBJECTIVES**

The objective of the proposed systematic review is to summarize the available evidence on the effects of prevention and intervention programs aimed at primary and secondary students for increasing school completion or reducing school dropout. Program effects on the closely related outcomes of school attendance (absences, truancy) will also be examined. Moreover, when accompanying dropout or attendance outcomes, effects on student engagement, academic performance, and school conduct will be considered.

The primary focus of the analysis will be the comparative effectiveness of different programs and program approaches in an effort to identify those that have the largest and most reliable effects on the respective school participation outcomes, especially with regard to differences associated with treatment modality, implementation quality, and program location or setting. In addition, evidence of differential effects for students with different characteristics will be explored, e.g., in relation to age or grade, gender, race/ethnicity, and risk factors. Because of large ethnic and socioeconomic differences in graduation rates, it will be particularly important to identify programs that may be more or less effective for disadvantaged students.

The ultimate objective of this systematic review is to provide school administrators and policymakers with an integrative summary of research evidence that is useful for guiding programmatic efforts to reduce school dropout and increase school completion.
METHODOLOGY

Criteria for inclusion and exclusion of studies in the review

Studies must meet the following eligibility criteria to be included in the systematic review.

Types of interventions
There must be a school-based or affiliated psychological, educational, or behavioral prevention or intervention program, broadly defined, that involves actions performed with the expectation that they will have beneficial effects on student recipients. School-based programs are those that are administered under the auspices of school authorities and delivered during school hours. School affiliated programs are those that are delivered with the collaboration of school authorities, possibly by other agents, e.g., community service providers, and which may take place before or after school hours and/or off the school grounds. Community-based programs that are explicitly presented as dropout prevention or intervention programs will be included whether or not a school affiliation is evident. Other community-based programs that may include dropout among their goals or intended outcomes, but for which dropout or related variables are not a main focus, and which have no evident school affiliation, will be excluded.

We expect that programs that might be excluded for being community-based with no school affiliation or dropout focus, but that happen to assess school dropout outcomes, would mainly be delinquency or drug prevention or treatment programs or. The rationale for this exclusion is that we believe these kinds of programs are likely to be outside the realm of strategies that school administrators might consider when selecting programs for dropout prevention or treatment.

Types of participants
The research must investigate outcomes for an intervention directed toward school-aged youth, defined as those expected to attend pre-k to 12th grade primary and secondary schools, or the equivalent in countries with a different grade structure, corresponding to approximately ages 4-18. The age or school participation of the sample must be presented in sufficient detail to allow reasonable inference that it meets this requirement. Recent dropouts who are between the ages of 18-21 will also be included if the program under study is explicitly oriented toward secondary school completion or the equivalent.

General population samples of school-age children will be included. Samples from populations broadly at risk because of economic disadvantage, individual risk variables, and closely related factors will also be included (e.g., inner city schools, students from low SES families, teen parents, students with poor attendance records, students who have low test scores or who are over-age for their grade).

Samples consisting exclusively of specialized populations, such as students with mental disabilities or other special needs, will not be included. The rationale for this decision is that dropout programs designed exclusively for students with mental or physical disabilities that generally prevent them for attending mainstream classes and typical schools are not likely to be applicable to mainstream students. However, inclusion of some such individuals in a broader sample in which they are a minority proportion does not make that broader sample ineligible. Students with learning disabilities, such as dyslexia, that generally don’t require them to be in
specialized schools or classrooms (i.e., they attend mainstream classes and typical schools) will be included.

**Types of research designs**
To be included a study must use an experimental or quasi-experimental design. Specifically, it must involve comparison of treatment and control conditions to which students are: (1) randomly assigned; (2) non-randomly assigned but matched on pretests, risk factors, and/or relevant demographic characteristics; or (3) non-randomly assigned but statistical controls (e.g., covariate-adjusted means) or sufficient information to permit calculation of pre-treatment effect sizes on key risk variables or student characteristics is provided. Treatment-treatment studies that compare two or more treatments to each other without a control group will be included if one treatment group receives a “sham” or “straw-man” treatment that is equivalent to a control condition, or if one of the treatments is a practice as usual condition in which that practice is not a distinctive program delivered at a relatively high level. Posttest-only non-equivalent comparisons (not randomized, matched, or demonstrating equivalence) will not be included. Single-group pretest-posttest designs will not be included.

**Types of outcomes**
To be included, a study must assess intervention effects on at least one eligible outcome variable. Qualifying outcome variables are those that fall in or are substantially similar to the following categories: (a) School completion/dropout; (b) GED completion/high school graduation; (c) Absences or truancy. If a measure absences, truancy, or attendance is the only outcome provided, the majority of the students in the sample must be age 12 or older. The rationale for this exclusion is practical; there is a large literature on programs designed to influence attendance for elementary school age children that is beyond the scope of this review. Moreover, there is already an active Campbell Collaboration protocol on this topic (Maynard, Tyson-McCrea, Pigott, & Kelly, 2009).

**Date and form of publications**
Eligible studies should be relatively modern to be applicable to contemporary students. Therefore, the date of publication or reporting of the study must be 1985 or later even though the research itself may have been conducted prior to 1985. If, however, there is evidence in the report that the research was actually conducted prior to 1980 (more than five years before the 1985 cutoff date), then the study will not be included. Eligible studies can be published in any language and conducted in any country as long as they meet all other eligibility criteria. Campbell Collaboration affiliates outside the United States will be asked to assist with the location of studies published in other countries and languages other than English.

**Study inclusion decision-making**
Inclusion and exclusion decisions will be based on readings of the full reports for each study judged potentially relevant during the search procedure (described below). Any questions or doubts about the inclusion of a study will be discussed with the primary reviewer and/or a second reviewer.

---

1 Note that there is no threshold for initial equivalence. To be eligible under the third criterion, studies must simply present statistically controlled data or information from which group equivalence effect sizes can be calculated for key risk factors or student characteristics such as age, gender, race/ethnicity, school attendance, school performance, etc. Should sufficient data be available, pretest effect sizes or pre-treatment effect sizes on risk factors or demographics may be used as covariates to adjust for initial treatment-control differences for all research designs.
Search strategy for identification of eligible studies

Resources searched
A comprehensive and diverse strategy will be used to search the international research literature for qualifying studies reported during the last 25 years (1985-2010). The wide range of resources searched is intended to reduce omission of any potentially relevant studies and to ensure adequate representation of both published and unpublished studies.

Electronic bibliographic databases to be searched include Dissertation Abstracts International, Education Abstracts, Education Resources Information Center (ERIC), ISI Web of Knowledge (Social Science Citation Index, SSCI), PsycINFO, and Sociological Abstracts.

Research registers to be searched include, the Cochrane Collaboration Library, the National Dropout Prevention Center/Network, the National Research Register (NRR), the National Technical Information Service (NTIS), and the System for Information on Grey Literature (OpenSIGLE). International research databases such as Australian Education Index, British Education Index, CBCA Education (Canada), Canadian Research Index will also be searched.

Reference lists in previous meta-analyses and reviews, and citations in research reports screened for eligibility will also be reviewed for potential relevance to the review. Correspondence with researchers in the field will also be maintained throughout the review process.

Search terms and keywords
A comprehensive list of search terms and key words related to the population, intervention, research design, and outcomes will be used to search the electronic bibliographic databases. These include the following terms, with synonyms and wildcards applied as appropriate:

School dropouts, school attendance, truancy, school graduation, high school graduates, school completion, GED, general education development, high school diploma, dropout, alternative education, alternative high school, career academy, schools-within-schools, schools and absence, chronic and absence, school enrollment, high school equivalency, school failure, high school reform, educational attainment, grade promotion, grade retention, school nonattendance, school engagement, and graduation rate;

AND

intervention, program evaluation, random, prevent, pilot project, youth program, counseling, guidance program, summative evaluation, RCT, clinical trial, quasi-experiment, treatment outcome, program effectiveness, treatment effectiveness, evaluation, experiment, social program, effective.

The following search terms will be used to exclude irrelevant studies: higher education, post-secondary, undergraduate, doctoral, prison, and inmate.

Description of methods used in primary research

Studies to be included in the review will employ experimental or quasi-experimental research designs that compare outcomes for an intervention group to those for a control or comparison condition. The control or comparison conditions in these studies include youth receiving no treatment, observation only, treatment as usual, or wait-listed control groups.
Most potentially eligible studies include both pretest and posttest measurements that allow calculation of pretest group equivalence, posttest group differences, and pretest-posttest changes. Pretest measurements generally occur at or immediately prior to the beginning of the prevention or intervention program, with posttest measurements occurring at or after the end of the program. The posttest measurements comparing the intervention and comparison conditions are the key outcome measurements of interest for the proposed review. Many studies measure outcomes at multiple follow-up points; the first follow-up occurring at or after program completion will be considered the posttest measurement and subsequent waves will be considered follow-up measurements.

One study that exemplifies the methods likely to meet the eligibility criteria for the proposed review is a program evaluation of Ohio’s Learning, Earning, and Parenting (LEAP) Program (Long et al., 1996). In 1989, almost 10,000 teenage parents throughout the state of Ohio were randomly assigned to the LEAP program or a no-treatment control group. The LEAP program used an incentive structure for teens to encourage regular attendance in a program designed to lead to a high school diploma or GED. Because it used random assignment, the LEAP study did not provide pretest group equivalence information for the intervention and control groups, as those differences were presumed negligible given the randomized study design (other studies using quasi-experimental designs, however, must provide such pretest information in order to be included in the proposed review). The key outcomes of interest from the LEAP study are the posttest measurements—in this case measured three years after random assignment. Outcomes measured at posttest included measures of the percent of students in the intervention and comparison conditions who completed 9th, 10th, and 11th grade, completed high school, completed a GED, or were currently enrolled in school or a GED program.

Criteria for determination of independent findings

Multiple reports from single studies, and multiple studies in single reports, will be identified through information on program details, sample sizes, authors, grant numbers and the like. If it is unclear whether reports and studies provide independent findings, the authors of the reports will be contacted.

All codable effect sizes will be extracted from study reports during the coding phase of the review (i.e., we plan to code multiple outcomes and multiple follow-ups measured within the same study). These will be separated according to the general constructs they represent (dropout, attendance, engagement, etc.) and each outcome construct category will be analyzed separately. We expect that some portion of the studies will provide more than one effect size for a particular outcome construct (e.g., report two measures of dropout). This circumstance creates statistical dependencies that violate the assumptions of standard meta-analysis methods. If there are relatively few instances of this for a given construct category, we will retain only one of these effect sizes in the analysis by selecting the construct that is most similar to those used by other studies in that category. For any construct categories where this is relatively common, however, we will retain all the effect sizes in the analysis and use the technique recently developed by Hedges, Tipton, and Johnson (2010) to estimate robust standard errors that account for the statistical dependencies.

---

2 For instance, constructs with similar measurement characteristics (source of information, length of measurement period, standardized assessment source).
Details of study coding categories

Eligible studies will be coded on variables related to study methods, the nature of the intervention and its implementation, the characteristics of the subject samples, the outcome variables and statistical findings, and contextual features such as setting, year of publication, and the like. A detailed coding manual is included in Appendix I. All coding will be done by trained coders who will enter data directly into a FileMaker Pro database using computer screens tailored to the coding items and with help links to the relevant sections of the coding manual. Effect size calculation is built into the data entry screens for the most common statistical representations and specialized computational programs and expert consultation will be used for the less common representations. We will select a 10% random sample of studies for independent double coding. The results will be compared for discrepancies that will then be resolved by further review of the respective study reports. The coding team will be retrained on any coding items that show discrepancies during this process.

Statistical procedures and conventions

Analysis will be conducted using SPSS and the specialized meta-analysis macros available for that program (Lipsey & Wilson, 2001) as well as Stata and the meta-analysis routines available for it (Sterne, 2009).

Effect size metrics

We anticipate using odds ratios as the effect size metric for dropout and other binary outcomes, and standardized mean difference effect sizes as the effect size metric for outcomes measured on a continuous scale (e.g., group differences in average attendance rates). All effect sizes will be coded such that larger effect sizes represent positive outcomes (e.g., less school dropout, higher attendance, less truancy).

All computations with odds ratios will be carried out with the natural logarithm of the odds ratios, defined as follows:

\[
\ln(OR) = \ln \frac{A \times D}{B \times C}
\]

where \(A\) and \(B\) are the respective counts of “successes” and “failures” in the treatment group, and \(C\) and \(D\) are the corresponding counts of “successes” and “failures” in the comparison group. The sampling variance of the logarithm of an odds ratio can be represented as:

\[
Var_{\ln(OR)} = \frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}
\]

Analytic results from the logged odds ratios effect sizes will be converted back to the original odds ratio metric for final substantive interpretation.
Standardized mean difference effect sizes \((d)\) will be used as the effect size metric for outcomes measured on a continuous scale:

\[
d = \frac{\bar{X}_{G2} - \bar{X}_{G1}}{s_p}
\]

where the numerator is the difference in group means for the intervention and comparison groups, and the denominator is the pooled standard deviation for the intervention and comparison groups. All standardized mean difference effect sizes will be adjusted with the small-sample correction factor to provide unbiased estimates of the effect size (Hedges, 1981). This small-sample corrected effect size \((g)\) and its sampling variance can be represented as:

\[
g = \left[1 - \left(\frac{3}{4N - 9}\right)\right] \times d
\]

\[
Var_g = \frac{n_{G1} + n_{G2}}{n_{G1}n_{G2}} + \frac{g^2}{2(n_{G1} + n_{G2})}
\]

where \(N\) is the total sample size for the intervention and comparison groups, \(d\) is the original standardized mean difference effect size, \(n_{G1}\) is the sample size for the intervention group, and \(n_{G2}\) is the sample size for the comparison group.

During the analytic phase of the project we will determine the number of coded effect sizes in the odds ratio and standardized mean difference metrics in each outcome construct category. If both occur in a given category, we will transform the effect size metric with the smaller proportion into the metric with the larger proportion using the Cox transform shown by Sánchez-Meca et al, (2003) to produce good results for this purpose. This will allow all the effect sizes for that outcome category to be analyzed together. If this involves a large proportion of the effect sizes in any category, sensitivity analyses will be conducted to ensure that the transformed effect sizes and those in the original metric produce comparable results.

**Missing data**

All reasonable attempts will be made to collect complete data on items listed in the coding manual (see Appendix I). Authors of the reports will be contacted if key variables of interest cannot be extracted from study reports. In the event that a small number of studies continue to have missing data on covariates or moderators of interest to be used in the final analysis, we plan to explore an option for imputing missing values using an expectation-maximization (EM) algorithm, which produces asymptotically unbiased estimates (Graham, Cumsille, & Elek-Fisk, 2003). A series of sensitivity analyses will be conducted to examine whether the inclusion of imputed data values substantively alters the results of the moderator analyses. If the EM algorithm fails to converge, or if other difficulties arise that make this technique not feasible, all resulting analyses will implement listwise deletion of missing data.

**Outliers**

The effect size distributions for each outcome construct category will be examined for outliers using Tukey’s (1977) inner fence as the criterion and any outliers found will be recoded to the inner fence value to ensure that they do not exercise disproportionate influence on the analysis results. The distribution of sample sizes will also be examined and any outliers similarly recoded.
to ensure that the corresponding weights are not excessively large in any analysis. For odds ratio effect sizes, this examination of outliers will be performed by examining the distribution of weights, rather than sample sizes.

**Analytic techniques**

All analysis with effect sizes will be inverse variance weighted using random effects statistical models. Specifically, the weighting function will be:

\[
    w_i = \frac{1}{Var_i + \tau^2}
\]

Where \( w_i \) is the weight for effect size \( i \), \( Var_i \) is the sampling variance for effect size \( i \) as defined above for the respective effect size metric, and \( \tau^2 \) is the random effects variance component estimated for each analysis with a method of moments or maximum likelihood estimator.

The unit of assignment to treatment and comparison groups will be coded for all studies, and appropriate adjustments will be made to effect sizes to correct for variation associated with cluster-level assignment (Hedges, 2007).

Summary and descriptive statistics of the study-level contextual characteristics, methodological quality characteristics, group and subject level characteristics, as well as outcome characteristics will be used to describe the eligible body of research studies. Main effects and moderator analysis will be conducted separately with the effect sizes in each outcome construct category with the latter done as multivariate (meta-regression) analysis when possible to minimize misleading results due to correlated independent variables. Random effects statistical models will be used throughout unless a compelling case arises for fixed effects analysis. Random effects weighted mean effect sizes will be calculated for all studies using 95% confidence intervals and displayed with forest plots. Estimates of Cochrane’s \( Q \), \( I^2 \), and \( \tau^2 \) will be used to assess variability in the effect sizes.

We plan to code pretest effect sizes when available. If available in sufficient numbers for certain outcomes, it may be possible to use the pretest effect sizes as covariates in our meta-regression models, to control for pre-treatment differences between treatment and comparison groups on the outcome variables. For the dropout outcomes, we do not expect to find many pretest effect sizes, because most programs are likely to involve students who are currently attending school (thus the pretest effect sizes would be zero). For attendance outcomes, we may have sufficient pretest effect sizes to use them in our analyses.

The main objective of the analyses, however, will be to describe the direction and magnitude of the effects of different interventions on the different outcome constructs in a manner that allows their comparative effectiveness to be assessed. Additionally, moderator analysis using meta-regression models will attempt to identify the characteristics of the interventions and student participants that are associated with larger and smaller effects for the various outcome constructs. Based on prior theory and research, the following moderators will be examined for their influence on effect sizes:

- Treatment modality
- Implementation quality
- Treatment duration
• Program location (typical school facilities, alternative schools, after school programs, non-school settings)\(^3\)
• Ethnic and or socioeconomic mix of sample
• Grade level of sample

Examination of funnel plots, the use of Duval and Tweedie’s trim and fill method (2000), and Egger’s regression test (1997) will be used to assess the possibility of publication bias and its impact on the findings of the review. Sensitivity analyses will be conducted to examine whether any decisions made during analyses substantively influenced the review findings, e.g., transformation between effect size metrics, the way outlier effect sizes and sample sizes were handled, the inclusion of studies with poorer methodological quality within the range allowed by the inclusion criteria, and missing data imputations.

**Treatment of qualitative research**

Qualitative research will not be included in this systematic review.

**SOURCES OF SUPPORT**

**External funding:**
Work on this review to date has been supported by a contract from the Campbell Collaboration.

**DECLARATIONS OF INTEREST**

The review authors have no conflicts of interest to report.

**AUTHOR(S) REVIEW TEAM**

**Lead reviewer:**
Sandra Jo Wilson, Ph.D.
Peabody Research Institute, Vanderbilt University
230 Appleton Place, PMB 181
Nashville, TN 37203-5721 USA
Phone: (615) 343-7215
Fax: (615) 322-0293
email: sandra.j.wilson@Vanderbilt.Edu

**Co-authors:**
1. Mark W. Lipsey, Ph.D. Director, Peabody Research Institute, Vanderbilt University.
2. Emily E. Tanner-Smith, Ph.D. Research Associate, Peabody Research Institute, Vanderbilt University.

---

\(^3\) Note that community-based programs without a school affiliation are not eligible for the review. As a result, we expect to find few (if any) programs that are not housed in typical or alternative school settings, or housed in school buildings but operate outside of school hours. Should we locate non-school based programs that are implemented under school auspices, we would include those in the moderator analysis.
3. Chiungjung Huang, Ph.D. Professor, National Changhua University of Education, Taiwan.

Content Expert Consultant:
Sandra L. Christenson, Ph.D.
Birkmaier Professor of Educational Leadership
University of Minnesota

ROLES AND RESPONSIBILITIES

Content: Wilson, Lipsey, Christenson
Systematic review methods: Wilson, Lipsey, Tanner-Smith
Statistical analysis: Wilson, Lipsey, Tanner-Smith, Huang
Information retrieval: Wilson, Tanner-Smith, Huang, Steinka-Fry

PRELIMINARY TIMEFRAME

Timeframe
The research team will begin working on the systematic review upon approval by the Campbell Collaboration editorial staff. We plan to complete the review by February, 2011.

REFERENCES


APPENDIX I: CODING MANUAL

STUDY LEVEL VARIABLES

Step 1. Study Identifiers, Study Context, Group Identification, and Study-Level Coding

STUDY IDENTIFIERS
The “unit” you will code here consists of a study, i.e., one research investigation of a defined subject sample or subsamples compared to each other, and the treatments, measures, and statistical analyses applied to them. Sometimes there are several different reports (e.g., journal articles) about a single study. In such cases, the coding should be done from the full set of relevant reports, using whichever report is best for each item to be coded; BE SURE YOU HAVE THE FULL SET OF RELEVANT REPORTS BEFORE BEGINNING TO CODE. Sometimes a single report describes more than one study, e.g., one journal article could describe a series of similar studies done at different sites. In these cases, each study should be coded separately as if each had been described in a separate report.

Each study has its own study identification number, or StudyID (e.g., 619). Each report also has an identification number (e.g., 619.01), which you will find printed on the folder holding the report. The ReportID has two parts; the part before the decimal is the StudyID, and the part after the decimal is used to distinguish the reports within a study. (These two types of ID numbers, along with bibliographic information, are assigned and tracked using the bibliography.) When coding, use the study ID (e.g., 619) to refer to the study as a whole, and use the appropriate report ID (e.g., 619.01) when referring to an individual report.

While reading reports for coding, be alert to any references to other dropout studies that may be appropriate to include in this meta-analysis. If you find appropriate-looking references that are not currently entered into the bibliography, the references may need to be entered.

[StudyID] Study identification number of the study you are coding, e.g., 1923.
[Coder] Coder’s initials (select from menu)
[CodeDate] Date you began coding this study (will be inserted automatically)

STUDY CONTEXT

[H1] Year of publication (four digits): If more than one report, choose earliest date.

[H2] Country in which study conducted.
   1. USA
   2. Great Britain
   3. Canada
   4. Scandinavia: Denmark, Finland, Norway, Sweden
   5. Australia/New Zealand
   6. Other Western European Country: __________
   7. Other: __________
Type of publication. If you are using more than one type of publication to code your study, choose the publication that supplies the effect sizes (in cases where more than one report provides effect sizes, choose a “peer reviewed” choice over another option, or choose the report that provides the most effect sizes).

1. Book
2. Journal article
4. Thesis or dissertation
5. Technical report
6. Conference paper/presentation
7. Other
8. Cannot tell

GROUP IDENTIFICATION AND SELECTION
At this stage, you will need to identify the aggregate treatment and/or comparison groups used in the study for which effect size statistics can be computed. To do this, you will need to distinguish aggregate groups, which you will code here, from subgroups (or breakouts), which you will code later:

(1) Aggregate treatment and/or comparison groups. The largest participant groupings on which contrasts between experimental conditions or contrasts between time points can be made. Note that the designations “comparison group” and “control group” refer to any group with which the treatment of interest is compared that is presumed to represent conditions in the absence of that treatment, whether a true random control or not. Often there is only one aggregate treatment group and one aggregate control group, but it is possible to have a design with numerous treatment variations (e.g., different levels) and control variations (e.g., placebos) all compared (e.g., in ANOVA format) to each other.

(2) Breakouts. Sometimes researchers will present data for some subset(s) of the participants from an aggregate group; e.g., for an aggregate group composed of males and females, the researchers may present some results for the males and females separately. You will code information about breakouts later.

Identifying the Aggregate Groups
Type in the name or identifier for each aggregate treatment group and each aggregate comparison group described in the study, whether you believe the group is eligible for coding or not.

Group labels used by researchers do not necessarily conform to the definitions of group types used in this project. In some cases, for example, researchers may compare one treatment with another treatment, and may call this “other” treatment a comparison or control group. For our purposes, if this “other” treatment group can realistically be expected to be effective, list it as a treatment group below; if it is a minimal or placebo treatment, not expected to produce an effect, list it as a comparison group.

Treatment Groups [H4a-d]

1
2
3
4
Comparison Groups [H5a-d]
1
2
3
4

[H4] Total number of treatment groups: _____
[H5] Total number of control groups: _____

ASSIGNMENT OF PARTICIPANTS

[H6] Unit of group assignment. The unit on which assignment to groups was based.
1. individual (i.e., some children assigned to treatment group, some to comparison group)
2. group (e.g., whole classrooms, schools, sites, facilities assigned to treatment and comparison groups)
3. program area, regions, school districts, counties, etc. (i.e., region assigned as an intact unit)
9. cannot tell

[H7] Method of group assignment. How participants/units were assigned to groups.
This item focuses on the initial method of assignment to groups, regardless of subsequent degradations due to attrition, refusal, etc. prior to treatment onset. These latter problems are coded elsewhere.
Random or quasi-random:
1. randomly after matching, yoking, stratification, blocking, etc. The entire sample is matched or blocked first, then assigned to treatment and comparison groups within pairs or blocks. This does not refer to blocking after treatment for the data analysis.
2. randomly without matching, etc. This also includes cases when every other person goes to the control group.
3. regression discontinuity design: quantitative cutting point defines groups on some continuum (this is rare).
4. quasi-random procedure presumed to produce comparable groups (no obvious differences). This applies to groups which have individuals apparently randomly assigned by some naturally occurring process, e.g. next person to walk in the door. The key here is that the procedure used to select groups doesn’t involve individual characteristics of persons so that the groups generated should be essentially equivalent.
Non-random, but matched or statistically controlled: Matching refers to the process by which comparison groups are generated by identifying individuals or groups that are comparable to the treatment group using various characteristics of the treatment group. Statistical control refers to inclusion of the matching variable as a covariate in an ANCOVA or multiple regression analysis. Matching can be done individually, e.g., by selecting a control subject for each intervention subject who is the same age, gender, and so forth, or on a group basis, e.g., by selecting comparison schools that have the same demographic makeup and academic profile of treatment schools. Similarly, statistical control variables can be used at either the individual or school level.
5. matched or statistically controlled ONLY on pretest measures of some or all variables used later as outcome measures.
6. matched or statistically controlled on pretest measures AND other personal characteristics, such as demographics.
7. matched or statistically controlled ONLY on demographics: big sociological variables like age, sex, ethnicity, SES.

Nonrandom, no matching prior to treatment but descriptive data, etc. regarding the nature of the group differences:

8. Non-random, not matched, but pretreatment equivalence information is available.

99. cannot tell

[H8] Confidence in assignment ratings. Overall confidence of judgment on how participants were assigned
1. Very Low (Little Basis)
2. Low (Best Estimate)
3. Moderate (Weak Inference)
4. High (Strong Inference)
5. Very High (Explicitly Stated)

Equivalence of the groups being compared
At this point, you should go to the Effect Size Database to code group equivalence effect sizes and descriptive information about initial group differences for the study. See the Effect Size Coding Sheet section of this manual for more information on effect size calculation.

[H9] Number of variables on which treatment and comparison group differences were statistically compared prior to the intervention. A statistical comparison is one in which a statistical test was performed by the authors, whether they provide data or not (e.g., “no statistically significant differences were found”). Do not include here any comparisons on pretest variables, that is, measures of a dependent variable taken prior to treatment, e.g., prior number of absences when subsequent number of absences is used as an outcome measure.

[H10] Results of statistical comparisons.
1. no comparisons made
2. no statistically significant differences
3. significant differences judged unimportant by coder. See note below regarding “importance” judgment.
4. significant differences, judged of uncertain importance by coder
5. significant differences, judged important by coder

[H11] Number of variables on which treatment and comparison group differences were or can be descriptively compared prior to the intervention. A descriptive comparison is any comparison across treatment and control groups that does not involve a statistical test (e.g., the actual number of males and females in each group or a statement by the author(s) about group similarity).

[H12] Results of descriptive comparisons.
1. no comparisons made or available
2. negligible differences, judged unimportant by coder. See note below regarding “importance” judgment.
3. some differences, judged of uncertain importance by coder
4. some differences, judged important by coder
Note: An “important” difference means a difference on several variables relevant to the outcome variables, or on a major variable, or large differences; major variables are those likely to be related to dropout, e.g., SES or poor academic performance.

[H13] Rating of similarity of treatment and control groups. Using all the available information, rate the overall similarity of the treatment group and the comparison group, prior to treatment, on factors likely to have to do with dropout or responsiveness to treatment (ignore differences on any irrelevant factors). Note: Greatest equivalence from “clean randomization” with prior blocking on relevant characteristics and no subsequent attrition/degradation; least equivalence with some differential selection of one “type” of individual vs. another on some variable likely to be relevant to dropout.

Guidelines: Use ratings in the 1-3 range for good randomizations and matchings, e.g., 1=clean random, 2=nice matched. Use ratings in the 5-7 range for selection with no matching or randomization or instances where it has been seriously degraded, e.g., by attrition before posttest. Within this bracket, the question is whether the selection bias is pertinent to the outcomes being examined. Were participants selected explicitly or implicitly on a variable that might make a big difference in dropout? The middle three points are for sloppy matching designs, degradations, bad wait list designs, and the like. If the data indicate equivalence but the assignment procedure was not random give it a 4 or thereabouts since not all possible variables were measured for equivalence between groups.

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

[H14] Overall confidence on rating of group similarity
1. Very Low (Little Basis)
2. Low (Best Estimate)
3. Moderate (Weak Inference)
4. High (Strong Inference)
5. Very High (Explicitly Stated)

[H15] Click here to record any problems you encountered while coding this section.

GROUP EQUIVALENCE EFFECT SIZE CODING

At this point, you should go to the Effect Size Database to code group equivalence effect sizes and descriptive information about initial group differences. See the Effect Size Coding section of this manual for more information on effect size calculation.

For each measure you can identify on which the treatment and control group were compared prior to treatment (other than dependent variables) or on which you can tell equivalence (e.g. if all males then code it here), determine which group is favored and if possible, calculate an effect size (ES, standardized difference between means or odds ratio). Do not include here any comparisons on pretest variables, that is, measures of a dependent variable taken prior to
treatment. In such cases the pretreatment ES is coded later as pretest information, not here as group equivalence information.

The only eligible variables for group equivalence effect sizes are: (a) gender, (b) age, (c) race/ethnicity, and (d) variables relating to risk for school dropout. A pretest that is used later in the study as a posttest would not be coded here – you would code it as a pretest effect size. If the study reports group equivalence outcome data for multiple risk variables, group equivalence effect size information should be coded for up to four variables. If more than four variables are available for any of the risk factors, code the four most relevant ones. When deciding which are most relevant, use the following criteria:

1. First preference should be given to behavioral measures (e.g., prior absences, school performance).
2. Second preference should be given to measures of psychological conditions, predispositions, or attitudes (e.g., school engagement, school bonding, etc.).
3. Lowest preference should be given to broad measures of social disadvantage or family history (e.g., socioeconomic status of parents, residence in inner-city).

[StudyID] Indicate the Study ID for the study you are coding.
[ReportID] Enter the Report ID for the report in which you found the information on group equivalence. Use the complete Report ID, e.g. 1973.01.
[pagenum] Enter the page number on which you found the information on group equivalence.

[ES24] Type of effect size:

5. Group Equivalence (for baseline treatment-control comparisons on variables other than the dependent variables)

[ES19] Wave number. Pretests and group equivalence effect sizes always get a 1; each wave thereafter gets numbered consecutively, beginning with 1. Some studies involve more than one posttest measurement and we need to be able to distinguish one from another. Give the first posttest after treatment a 1, the second a 2, and so on.

[ES15] Variable on which comparison is made: ____________________________ (e.g., gender, age, etc.)

[ES17] Which group is favored? Whichever group has more of the characteristic that presumably makes them better off or more amenable to treatment (e.g., less truant, higher SES, smarter, etc.) is considered favored. NOTE: You should code this item even for cases in which you are unable to calculate a numeric effect size but have information about which group is favored.

1. Treatment (fewer males, younger, fewer minorities, less risk)
2. Control (fewer males, younger, fewer minorities, less risk)
3. Neither, exactly equal
9. Cannot tell, no report

Data Fields: Fill in the data fields using the relevant statistical information provided in the report(s). You do not need to fill in all the fields; fill in only the information necessary to calculate an effect size. Thus, if the report provides sample sizes, means, standard deviations, and t-test scores, you need only enter the sample sizes, means, and standard deviations.
ONCE YOU HAVE FINISHED CODING THE GROUP EQUIVALENCE EFFECT SIZE INFORMATION, YOU SHOULD RETURN TO THE STUDY LEVEL FILE TO COMPLETE THE CODING OF FOR THAT SECTION.

TREATMENT AND CONTROL GROUPS CODING

Create one record in this database for each of the aggregate treatment and/or control groups that you selected earlier for coding. Studies with a treatment group and a control group will have two records, etc.

Group Identification and General Nature of Treatment

[StudyID] Type in the StudyID for the study you are coding if it does not appear automatically.

[GroupID] Number each group consecutively within a study, starting with 1.

[G1] Select the type of group you are coding.
   1. Treatment group
   2. Control group

[G2] What general type of “treatment” does this group receive?

Intervention Condition

1. Focal program or treatment. There may be several focal programs in a study, as when two different types of treatments, both of which could be expected to be effective, are compared.

Control Condition

2. “Straw man” alternate program or treatment, diluted version, less extensive program, etc., not expected to be effective but used as contrast for treatment group of primary interest. If the alternate treatment is not minimal and could realistically be expected to be effective, it is not a control condition and should be classified as a focal treatment instead.

3. Placebo (or attention) treatment. Group gets some attention or sham treatment (e.g., watching Wild Kingdom videos while treatment group gets therapy)


5. No treatment. Group gets no treatment at all. Note: The difference between “no treatment” and “treatment as usual” hinges on whether or not the treatment and control groups in this study have an institutional framework or experience in common.

[G3] Program name. Write in program or treatment label for this group (e.g., Dropout Prevention Curriculum, waiting list control, etc.). REMEMBER: YOU MUST CREATE A PROGRAM LABEL FOR CONTROL GROUPS AS WELL AS TREATMENT GROUPS.

[G4] Program description. Write in a brief description of the treatment this group receives. Please try to keep the description short by focusing on the key elements of treatment, but make sure you include ALL treatment elements in your description. As much as possible, quote or give a close paraphrase of the relevant descriptive text in the study report. REMEMBER: YOU MUST CREATE A DESCRIPTION FOR CONTROL GROUPS AS WELL AS TREATMENT GROUPS.
TREATMENT CHARACTERISTICS

[G5] Intervention type:
1. School-based (administered under the auspices of school authorities and delivered during school hours)
2. School affiliated (delivered with the collaboration of school authorities, possibly by other agents, e.g., community service providers; may take place before or after school hours and/or off the school grounds)
3. Community-based (explicitly presented as dropout prevention/intervention programs; may or may not have a school affiliation)
4. Not applicable (control condition)


For each treatment AND control condition:
First check all program types that apply to a given intervention (e.g., a program may include GED preparation, tutoring, and contingency management).

Second, choose the one program type that can be considered the focal program characteristic. Most programs will arguably deliver multiple service types, but do your best to narrow the focal type down to one category. It may be helpful to examine the amount of each service type delivered. For instance, if a program delivered 1 hour/week of skills training to parents and 5 hours/week of vocational training to students, you would code vocational training as the focal program component. If a program contains too many service types to distinguish a focal type, choose “multi-service” package as the focal component.

ACADEMIC:
1. Curriculum
2. Academic program
3. Remedial education (e.g., reading remediation)
4. GED preparation
5. Computer-assisted learning
6. Test-taking and study skills assistance
7. Tutoring
8. Homework assistance
9. Extracurricular activities (e.g., after school club). NOTE: just because a program is delivered after school does not mean it should be coded here; this program component should include academic, social, or sport activities that are separate from regular school activities.
10. Professional development for school staff

SCHOOL STRUCTURE
11. Class or grade reorganization (schools within schools)
12. Small class sizes/small “learning communities”
13. Alternative school

FAMILY ENGAGEMENT:
14. Family outreach
15. Feedback to parents and students on performance
16. Parent or teacher consultation enhancement
17. Parenting skills program
COLLEGE FOCUSED/CONNECTING STUDENTS TO ATTAINABLE FUTURE:
18. Academic advising
19. College-preparatory curriculum
20. Academic summer program
21. College campus visits
22. College and financial aid application assistance
23. College scholarships

WORK RELATED/ FINANCIAL SUPPORT:
24. Internships
25. Career exploration
26. Vocational training
27. Job placement assistance
28. Living allowance
29. Bonuses and sanctions applied to welfare grant

LINKING TO SERVICES:
30. Case management
31. Health services
32. Transportation assistance
33. Child care/day care
34. Residential living services

SOCIAL RELATIONSHIPS:
35. Mentoring
36. Peer support
37. Social events
38. Community service/volunteer service/tutoring (“helper-therapy”)
39. Recreational, wilderness, etc. program

PERSONAL/AFFECTIVE:
40. Counseling
41. Skills training (life skills, social skills/social competence)
42. Cognitive behavioral therapy (e.g., problem solving skills)

BEHAVIORAL:
43. Attendance monitoring
44. Contingency management, financial incentives, token economy, extrinsic reward system to promote attendance/academic achievement

OTHER:
45. Multi-service package (NOTE: Only choose this program code if the group receives an amorphous, broadly defined program with components that cannot be clearly identified otherwise. Use this program code as focal if a group has multiple “focal” treatment components and you cannot make a distinction otherwise.

46. OTHER (Please, describe [prog50a]_____________)
88. Control group
[G9] **Treatment Site.** Nature of the site in which treatment generally delivered: (select one)

**School Sites**
1. Regular Class Time (this includes interventions delivered during regularly scheduled classes AND in the regular classroom for youths in the group)
2. Special Class (e.g., youth in treatment are in a classroom-type setting that is different from a typical classroom, although it may be the subjects’ usual classroom – includes such settings as special education classrooms, schools-within-schools, alternative schools, etc.)
3. Resource Room, School Counselor’s Office, or other similar setting that is NOT the student’s regular classroom; the idea here is that students are removed from class for treatment
4. Treatment delivered at school facility, but not during regular school hours (e.g., afterschool programs)

**Home**
9. Treatment delivered in the subject’s home

**Community-based, Non-residential**
10. Private office, clinic, center (e.g., YMCA, university, therapist’s office)
11. Public office, clinic, center (e.g., human services department, public health agency)
12. Work site (e.g., community service, trash collection on roadside, etc.)
13. Park, playground, wilderness area, etc.

**Institutional, Residential**
14. Private institution, residential
15. Public institution, residential (e.g., camp, reformatory)

**Mixed or Multiple Sites**
16. School and home
17. Other mixed, some combination of above sites (NOTE: if all sites are school based, use 16 above)

88 N/A: control group
99 Cannot tell

[G10] **Role of the evaluator(s)/author(s)/research team or staff in the program.** This item focuses on the role of the research team working on the evaluation, regardless of whether they are all listed as authors.

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

[G11] **Role of program developer in the research project.** This item focuses on the individual (or group of individuals) who created or developed the program and their role in the delivery of the program under study. Is the program developer the researcher conducting the study, or is the program developer not participating in the research project?

1. Program developer is author/evaluator/delivery agent
2. Delivery agent/author/evaluator modified existing program, but original program developer is not involved (note: this response suggests that the author/evaluator/delivery agent takes on a sort of quasi-developer status by modifying a program)

3. Program developer is not affiliated with research study and program is delivered as originally intended by developer

**[G12]** Routine practice or program vs. research project. Indicate the appropriate level for the treatment you are coding: at one end of the continuum are research projects (option 1), in which a researcher decides to implement and evaluate a particular program for research purposes; in many cases, the program may require the cooperation of a service agency (school, clinic, etc.), but the intervention is delivered primarily so the researcher can conduct research. At the other end of the continuum are evaluations of “real-world” or routine programs (option 3): a service agency implements a program on its own, and also decides to conduct an evaluation of the program; the evaluation may or may not be conducted by outside researchers. In the middle of the continuum are demonstration projects (option 2), which are conducted primarily for research purposes, but generally have more elements of “real world” practice than typical research projects as defined under option 1. Demonstration projects generally involve a program that has been studied in prior research but is being tested for effectiveness in different settings than the original research, or on a larger scale than the original research.

If a researcher is a school principal or some other school staff person and is conducting the evaluation as part of his/her dissertation, the decision depends on the extent of the program. If the program is small-scale and implemented in, say, a classroom or two, and supervised by the researcher/principal, code it as a research project. If the program is a broader school-wide program that the researcher/principal happens to be evaluating, code it as either a demonstration or routine program, depending on whether the program is a special program being tested (demonstration) or something that the school does on a routine basis (routine practice).

1. research project: The intervention would not have been implemented without the interest or initiative of the researcher(s). The intervention is delivered by the research staff or by service providers (regular agency personnel, teachers, etc.) trained by the researchers.

2. demonstration project: A research project that involves a new or special program being tested, rather than a routine program. Although generally implemented by researchers for research purposes, a demonstration project has more elements of actual practice than a research project. Demonstration projects usually involve programs that have been studied previously, either in small-scale pilot projects or tightly controlled efficacy trials; demonstration projects would serve as a larger scale or quasi-real-world test of a promising program.

3. evaluation of a “real-world” or routine program: A service agency implemented the program using routine personnel and the typical clients for that program; there may be outside researchers who conduct the evaluation, but the program they are evaluating was already in place before the research began and is presumed to continue after the research has ended.
[G13] Treatment provider’s discipline. Indicate the discipline or type of treatment provider for the treatment. This item focuses on the individual(s) who have direct contact with the subjects in treatment, not necessarily the persons conducting the data analysis or evaluation.

1. Teacher
2. School guidance counselor
3. School psychologist
4. School personnel, other than school counselor or teacher (e.g., principal, school nurse)
5. Counselor
6. Social worker
7. Researcher or researcher’s staff, graduate students
8. Other
88. N/A: no treatment received
99. Cannot tell

[G14] Did treatment personnel receive special training in this specific program, intervention, or therapy? If the treatment is delivered by the researcher, use “yes” below, unless the report indicates otherwise.

1. yes
2. no
9. cannot tell

[G15] If yes, write in amount of training of personnel for providing this treatment:

________________________

[G16] Treatment Format:

For each treatment AND control condition:

First check all formats that apply to a given intervention (e.g., a program may include group and individual components, or have a family component).

Second, choose the one format type that can be considered the focal format. This selection should match the format of the focal program type you selected above under G6. If you selected multi-service package above, select the format for the most frequent or most focal piece of the package; if this is impossible, select multiple format program.

1. Subject alone (self-administered treatment)
2. Subject & provider, one-on-one
3. Subject group and provider, not classroom
4. Subject group and provider, classroom
5. Parents only and provider, child not present
6. Group of parents and provider, children not present
7. Child & parents with provider
8. Group of families with provider
9. Child & parents, no provider (self-administered treatment)
10. Teachers, treatment professional, no children
11. Parents alone (self-directed)
12. Multiple format program; no focal format
88. N/A: control group
Focal Treatment Implementation/Length/Integrity

[G20] Duration of treatment. Approximate (or exact) number of weeks that subjects received treatment, from first treatment event to last excluding follow-ups designated as such. Divide days by 7; multiply months by 4.3. Code 888 if a control group that receives nothing. Code 999 if cannot tell. Estimate for this item if necessary, and if you can come up with a reasonable order of magnitude number.

[G22] Approximate (or exact) frequency of contact between subjects and provider or treatment activity. This refers only to the element of treatment that is different from what the control group receives.
   1. less than weekly
   2. Once a week
   3. 2 times a week
   4. 3-4 times a week
   5. daily contact (not 24 hours of contact per day but some treatment during each day, perhaps excluding weekends)
   6. continuous (e.g. residential living)
   9. cannot tell
   88. N/A: control group

[G24] ____________ Approximate (or exact) mean hours actual contact time between subject and provider or treatment activity per week if reported or calculable. Assume that high school classes, counseling, or therapy sessions are an hour unless otherwise specified. Round to one decimal place. Code 8888 for institutional, residential, or around the clock program; code 9999 if not available.

[G26] ____________ Approximate (or exact) mean number of hours total contact between subject and provider or treatment activity over full duration of treatment per subject if reported or calculable. Round to whole number. Code 8888 for institutional, residential, or around the clock program; code 9999 if not available.

[G29] Monitored treatment implementation. Was the implementation of the program monitored by the author/researcher or program personnel to assess whether it was delivered as intended?
   1. Yes. Do not infer that monitoring happened. Select “yes” ONLY if the report specifically indicates that implementation was monitored.
   0. No
   9. Cannot Tell

[G30] Based on evidence or author acknowledgment, was there any uncontrolled variation or degradation in implementation or delivery of treatment, e.g., high dropouts, erratic attendance, treatment not delivered as intended, wide differences between settings or individual providers, etc.? Assume that there is no problem if one is not specified.

This question has to do with variation in treatment delivery, not research contact. That is, there is no “dropout” if all subjects complete treatment, even if some fail to complete the outcome measures.
   1. yes (describe below)
   2. possible (describe below)
   3. no, apparently implemented as intended
Implementation Problems. Click to describe implementation problems, if any.

Subject Characteristics

Gender composition of group.
1. no males (<5%)
2. some males (<50%)
3. 50% to 60% male
4. mostly males (>60%)
5. all males (>95%)
6. cannot tell

Enter percent male: __________

ETHNICITY CODING:

Percent white
Percent black
Percent Hispanic
Percent other minority
Percent non-white (ONLY use this category if specific minority groups are not mentioned; if you use this category, there should only be numbers in the white and non-white categories)

Rankings: 1=clear majority; 2=present but proportion unknown; 3=clear minority; 0=not present.

White rank
Black rank
Hispanic rank
Other minority rank
Non-white rank (ONLY use this category if specific minority groups not mentioned; if you use this category, there should only be numbers in the white and non-white categories)

Describe others and/or non-whites: ____________________________.

Enter the average age of the sample using number of years.

High and low age using years.

Enter the average grade level of the sample. (dropdown menu)

High and low grades (dropdown menu)

Predominant level of “risk” of youths in the sample:
____________________________. Think of the reason that the subjects in this group ended up in this group; did the researchers select potential dropouts for treatment; if yes, how were the potential dropouts identified?

Socioeconomic status: Type in a brief description of the socioeconomic composition of the sample. This might include information on the percentage of children eligible for free lunches, the income level of the children’s parents, or a description of poverty in the community. Quote or closely paraphrase the relevant descriptive information in the report.
DEPENDENT VARIABLES CODING

Select the general construct group for the dependent variable you are coding, then select the specific construct category that best matches the dependent variable.

[DV1] Construct Group

100. Dropout
101. Attendance, truancy
102. Academic performance
103. School conduct
104. School engagement

[DV2] Specific Construct

Dropout
200. Dropout
201. Graduation
202. GED completion
203. Enrolled in post-secondary education

Attendance
204. Absences/truancy
205. Tardies
206. Attendance

Academic performance
207. GPA, grades
208. Standardized test scores
209. Academic track
210. Grade retention
211. Unstandardized, generic academic achievement score

School conduct
212. Suspensions
213. Expulsions
214. Detention
215. Classroom behavior

School engagement
216. School self-concept
217. Academic expectations/goal setting
218. Attitude toward school/school bonding
219. Attitude toward teachers

[DV3] Source of information. Who provided the information for this dependent variable?
1. Participants, self-report
2. Parents
3. Peers
4. Teachers
5. Principal
6. Service Provider (treatment agent)
7. School Records
8. Researcher or interviewer
9. Involved other (not treatment agent, not researcher), e.g., school counselor.
10. Multiple sources, cannot tell which is dominant
99. Cannot tell

[DV4] Type of Measure.
1. Survey, questionnaire, or interview
2. Standardized test (e.g., standardized achievement test)
3. School records
4. Other: __________
9. Cannot Tell

BREAKOUT/SUBGROUP CODING

Breakouts are comparisons involving subgroups of an aggregate treatment and/or control group. For example, the males in a treatment group might be compared with the males in a comparison group, or pretest-posttest results might be presented for males and females separately. Each variable (e.g., gender, age) by which the aggregate group(s) are subdivided constitutes one breakout, and each value of that variable defines one subgroup; i.e., a males vs. females stratification is one breakout (gender) with two subgroups, one male and one female. If only the male subgroup is reported, there is still one breakout, but only one subgroup.

Note that a simple report of the number of males and females in the treatment and control groups does not constitute a breakout (though it is relevant to group equivalence issues). To be a breakout, outcome data must be reported for the treatment-control or pretest-posttest comparison for at least one subgroup of the breakout variable. Breakouts are usually presented because the authors think that subgroups (e.g., males and females) are sufficiently different to warrant separate presentation of results (because, for example, males may be more likely to dropout than females).

NOTE: Only certain breakout variables are eligible for coding. These include gender, age, ethnicity, and prior school completion/dropout, GED completion, or absences/truancy. If you encounter another breakout variable that may be relevant to dropout, please check with Sandra. Create a new record for each subgroup that you will be coding for this study.

[StudyID] Study ID for the study you are coding.

[BreakID] Subgroup number. Assign a number to the subgroup such that the first subgroup you code is numbered 1, the second is numbered 2, and so on. These numbers are used within a study, so when you code subgroups from another study, you would start over with 1 again.

[Labels:B2] Write in descriptor for the subgroup you are coding, e.g., males, 8 year olds, whites, etc.

EFFECT SIZE CODING

Step 1. General Information

[StudyID] Type in the appropriate StudyID if it does not appear automatically.
[ReportID] Report ID for this effect size. Indicate the report number (e.g., 2098.01) for the report in which you found the information for this effect size. This is important so that we can find the source information for the effect sizes later on, if necessary, and is especially important for studies with multiple reports.

[ESID] Effect size ID. FileMaker will automatically generate unique effect size ID numbers ACROSS studies.

[pagenum] Page number for this effect size. Indicate the page number of the report identified above on which you found the effect size data. If you used data from two different pages, you can type in both, but use a comma or dash between the page numbers.

There are 3 types of effect sizes that can be coded: pretest, posttest, and group equivalence (or baseline similarity) effect sizes. They are defined as follows:

- **Pretest effect size.** This effect size measures the difference between a treatment and comparison group before treatment (or at the beginning of treatment) on the same variable used as an outcome measure, e.g., school attendance measured before the treatment begins is used as a pretest for school attendance measured the same way after the treatment ends.

- **Group equivalence effect size.** Group equivalence effect sizes are used to code the equivalence of two groups prior to treatment delivery on variables that might be related to outcome. See the Group Equivalence Coding section for more information.

- **Posttest effect size.** This effect size measures the difference between two groups after treatment on some outcome variable.

This is very important!!!! These three types of effect sizes are different from the multiple breakouts and multiple dependent variables that you might have in a study. For example, you might have a study that measures the treatment and comparison groups at pretest and posttest at 6 months after treatment on 3 different dependent variables. The results might be presented for the entire sample and broken down by gender. In this case you would have 6 group comparison effect sizes for the entire sample – three for the pretest and 3 for the 6 month posttest (the three is for your three dependent variables). In addition to these 6 aggregate effect sizes, you will have 6 more for the girls (the same as for the aggregate groups but just for the subgroup of girls) and 6 for the boys (also the same as for the aggregate groups but just for the subgroup of boys).

[ES24] Type of effect size:

1. Pretest (for treatment-control baseline comparison on a dependent variable)
2. Posttest (for the first treatment-control outcome comparison on a dependent variable)
3. Group Equivalence (for baseline treatment-control comparisons on variables other than the dependent variables)

[ES19] Wave number. Pretests and group equivalence effect sizes always get a 1; each wave thereafter gets numbered consecutively, beginning with 1. Some studies involve more than one posttest measurement and we need to be able to distinguish one from another. Give the first posttest after treatment a 1, the second a 2, and so on.

[ES47] Timing of measurement. Approximate (or exact) number of weeks after treatment when measure was taken. Divide days by 7; multiply months by 4.3. Enter 999 if cannot tell, but try to make an estimate if possible. Enter 0 if pretest. [es47_ck]
Step 2. Group Selection

[GroupID1] Group 1: Treatment group

[GroupID2] Group 2: Control group

[BreakID] Select Breakout group if relevant.

Step 3. Dependent Variable Selection

[VarNo] Select the dependent variable for this effect size.

Step 4. Effect Size Calculation and Data Entry

It is now time to identify the data you will use to calculate the effect size and to calculate the effect size yourself if necessary (see below). Effect sizes can be calculated ONLY from data based on the number of subjects, e.g., average number of days absent per subject and the corresponding standard deviation) or proportion of subjects who were chronic truants during a given time period. Effect sizes can NOT be calculated from data based solely on the incidence of events, e.g., total number of days absent per group. THIS IS VERY IMPORTANT—BE SURE YOU KNOW WHICH KIND OF DATA YOU HAVE.

You need to determine what effect size format you will use for each effect size calculation. There are two general formats you can use, each with its own section in FileMaker:

1. Compute ES from means, sds, variances, test statistics, etc.
2. Compute ES from frequencies, proportions, contingency tables, odds, odds ratios, etc.

Also note that within each of the above effect size formats, effect sizes can be calculated from a variety of statistical estimates; to determine which data you should use for effect size calculation, please refer to the following guidelines in order of preference:

1. Compute ES from descriptive statistics if possible (means, sds, frequencies, proportions).
2. If adequate descriptive statistics are unavailable, compute ES from significant test statistics if possible (values of t, F, Chi square, etc.).
3. If significance tests statistics are unavailable or unusable but p value and degrees of freedom (df) are available, determine the corresponding value of the test statistic (e.g., t, chi-square) and compute ES as if that value had been reported.

Note that if the authors present both covariate adjusted and unadjusted means, you should use the covariate adjusted ones. If adjusted standard deviations are presented, however, they should not be used.

[ES17] Which group is favored?
Select the group that has done “better”:

1. Treatment
2. Control
3. Neither, Exactly Equal
4. Cannot tell

For treatment-control comparisons, the treatment group is favored when it does “better” than the control group. The control group is favored when it does “better” than the treatment group. Remember that you cannot rely on simple numerical values to determine which group is better off. For example, a researcher might assess the attendance and report this variable in terms of
the average number of absences in the last semester. Fewer absences are better than more, so in this case a lower number, rather than a higher one, indicates a more favorable outcome.

Sometimes it may be difficult to tell which group is better off because a study uses multi-item measures in which it is unclear whether a high score or a low score is more favorable. In these situations, a thorough reading of the text from the results and discussion sections usually can bring to light the direction of effect – e.g., the authors will often state verbally which group did better on the measure you are coding, even when it is not clear in the data table. Note that if you cannot determine which group has done better, you will not be able to calculate a numeric effect size. (You will still be able to create an effect size record—just not a numeric effect size.)

[ES23] Effect size derived from what type of statistics?
1. Means and SDs; means and variances; means and standard errors
2. N successful/unsuccesful (frequencies)
3. Proportion successful/unsuccesful (percentage successful or not)
4. Multi-category (polychotomous) frequency or %
5. Independent t-test
6. Probability (p-value) with N or degrees of freedom
7. One-way ANOVA (2 groups, 1 degree of freedom)
8. One-way ANOVA (>2 groups, >1 degree of freedom)
9. Factorial Design (Repeated measures ANOVA, 2x2 ANOVA, MANOVA, etc.)
10. Covariance Adjusted (ANCOVA)
11. Chi-square statistic (1 degree of freedom; from 2x2 table)
12. Chi-square (from larger than 2x2 table)
13. Nonparametric statistics (Mann Whitney, etc.)
14. Correlation coefficient (zero-order)
15. Effect sizes as reported directly in the study
16. Other (please specify)

[ES50] For this effect size, did you use adjusted data (e.g., covariate adjusted means) or unadjusted data? If both unadjusted and adjusted data are presented, you should use the adjusted data for the group means or mean difference, but use unadjusted standard deviations or variances. Adjusted data are most frequently presented as part of an analysis of covariance (ANCOVA). The covariate is often either the pretest or some personal characteristic such as socioeconomic status. If you encounter data that is adjusted using something other than a covariate, please see Sandra or Mark.
1. Unadjusted data
2. Pretest adjusted data (or other baseline measure of an outcome variable construct)
3. Data adjusted on some variable other than the pretest (e.g., socioeconomic status)
4. Data adjusted on pretest plus some other variables

[ES22] Confidence in effect size calculation
1. High Estimate (e.g., have N and crude p values only, e.g., p<.10, and must reconstruct via rough t-test equivalence)
2. Moderately Estimated (e.g., have complex but relatively complete statistics, e.g., multiple regression, multifactor ANOVA, etc. as basis for estimation)
3. Some Estimation (e.g., have unconventional statistics and must convert to equivalent t-values or have conventional statistics but incomplete, such as exact p values only)
4. Slight Estimation (e.g., must use significance testing statistics rather than descriptive statistics, but have complete statistics of the conventional sort, such as a t-value or F-value)
5. No Estimate (e.g., have descriptive data: means, sds, frequencies, proportions, etc.; can calculate an ES directly.)

[ES44] Significance information for this comparison.
Did the authors provide any information about the statistical significance of the difference between the two groups you selected on the dependent variable you selected for the time point you have selected for this comparison? Sometimes authors will state that a particular comparison was not significant, but not provide any calculable effect size data. In these cases, you should select “5” for this item. The effect size field should remain blank. In other cases, authors will state that a particular comparison was significant, but not provide any calculable effect size data. In these cases, you should select “4” for this item. Again, the effect size field should remain blank. NOTE: the last three options (4, 5, and 6) are for cases for which you have direction (i.e., you know which group is favored) but no effect size information.

1. Significant result, ES data below
2. Non-significant result, ES data below
3. Significance not reported, ES data below
4. Significant result, no ES data
5. Non-significant result, no ES data
6. Significance not reported, no ES data

[ES55] Intent-to-treat analysis: Are results for this effect size based on an intent-to-treat analysis?
Experimental and quasi-experimental designs may employ “intent-to-treat” (ITT) or “completer” analyses. An intent-to-treat analysis is one that (attempts to) include outcome data from all the participants initially assigned to the treatment and comparison conditions regardless of their compliance with the entry criteria, the treatment they actually received, or any subsequent withdrawal from treatment (non-completers) or deviation from the protocol. A true ITT is possible only when the authors (attempt to) use outcome data for all randomized (or otherwise assigned) subjects; if all assigned subjects are used to present outcome results, then code as ITT, regardless of whether authors call the analysis an ITT. If the authors attempt to collect outcome data on non-completers and even if they are not 100% successful in this attempt, still code as ITT (as the missing data for non-completers is then coded as attrition). Sometimes researchers will use a modified ITT, in which they estimate missing data on non-completers, or include all subjects with pretests but not all who were randomized. These modified ITTs would be coded as “2” below. Completer analyses (AKA ‘treatment on the treated (TOT)’ analyses) involve only the participants who completed treatment or met some other criteria indicating an acceptable level of participation.

1. Intent-to-treat analysis (all subjects who were assigned are used in posttest)
2. Modified intent-to-treat (not all assigned subjects are used in posttest, but authors have done some modifications to approximate a true ITT)
3. Completer analysis (only those subjects who completed treatment or who stayed in the study are used in posttest)

[ES15] Significance of group equivalence comparison (ONLY).

1. No statistically significant differences
2. Statistically significant differences
3. Negligible descriptive differences
4. Significant descriptive differences
98. N/A: No comparison made
Assigned and Observed N
Assigned N, Observed N. These fields refer to the number of subjects who were originally assigned to the group(s) involved in this effect size (Assigned N) and to the number of subjects who were actually “observed” or “measured” (Observed N). If you cannot tell how many subjects were originally assigned to a group, look at the number of subjects (Observed N) at pretest; you can frequently use pretest sample sizes for assigned N. However, in cases where the authors have removed the subjects who do not have both pretest and posttest measures (such that the pretest N and the posttest N are the same), do not assume that the number of subjects at pretest is the correct number for Assigned N and, instead, leave this field blank. In cases where there is no attrition, the Assigned N is the same as the Observed N. Only use the same numbers for Assigned N and Observed N when you are SURE that there is no attrition.

[ES36] Assigned N for the treatment group (or pretest, if this is a pretest-posttest effect size).
[ES37] Assigned N for the comparison or second treatment group (or posttest, if this is a pretest-posttest effect size; if this is a pretest-posttest effect size, this value should be the same as the assigned N for the pretest).
[ES38] Total Assigned N.
[ES1] Observed N for the treatment group (or pretest, if this is a pretest-posttest effect size).
[ES2] Observed N for the comparison or second treatment group (or posttest, if this is a pretest-posttest effect size).
[ES3] Total Observed N.

[ES51] Number of units assigned for treatment group (for cluster-assigned studies): _____
[ES52] Number of units assigned for control group (for cluster-assigned studies): _____
[ES53] Intra-class correlation (ICC) for outcome measure (for cluster-assigned studies): _____

Other Effect Size Data Fields
[ES9] Mean for treatment group
[ES10] Mean for comparison group
[ES11] Difference in group means
[ES12] Standard deviation for treatment group
[ES13] Standard deviation for comparison group
[ES14] Pooled sd
[ES31] N successful for treatment group
[ES32] N successful for comparison group
[ES33] N failed for treatment group
[ES34] N failed for comparison group
[ES4] Dependent t-value
[ES5] Independent t-value
[ES6] $\chi^2$ (df=1)
[ES20] Effect size reported by authors
[ES60] Odds ratio reported by authors

Final Effect Size Determination
[ES21] Effect size value- standardized mean difference
[ES81] Effect size value- odds ratio

Remember that you cannot rely on simple numerical values to determine which group has done better. For treatment-control comparisons, a positive effect size should indicate that the treatment group did “better” on the outcome measure than the comparison group, while a...
negative effect size indicates that the comparison group did “better” than the treatment group, and a zero effect size means that the two groups are exactly equal on the measure. You must make sure that the sign of the effect size matches the way we think about direction, such that the effect size is positive when the treatment group is better and negative when the comparison group is better.

Effect sizes can range anywhere from around –3 to +3. However, you will most commonly see effect sizes in the –1 to +1 range. Odds ratios smaller than 1 indicate that the control group is better off; those greater than 1 indicate that the treatment group has the better outcome.

Note: If the authors report an effect size, include that in your coding and use it for the final effect size value if no other information is reported. However, if the authors also include enough information to calculate the effect size, always calculate your own and report it in addition to that reported in the study.

[ES39] Any problems coding this effect size?