TASK FORCE ON EVIDENCE-BASED INTERVENTIONS IN SCHOOL PSYCHOLOGY

Sponsored by: Division 16 of the American Psychological Association

and

The Society for the Study of School Psychology

Endorsed by: The National Association of School Psychologists

For further information contact:

Thomas R. Kratochwill, PhD
Chair, Task Force on Evidence-Based Interventions in School Psychology
School Psychology Program
1025 W. Johnson Street
University of Wisconsin-Madison
Madison, WI 53706
tomkat@education.wisc.edu
(608) 262-5912
Members and Domains of the Task Force on Evidence-Based Interventions in School Psychology*

Task Force, Chair
Thomas R. Kratochwill

<table>
<thead>
<tr>
<th>Conceptual and Methodological Issues</th>
<th>Liaison Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lynn Fuchs</td>
<td>Dick Abidin (Division 53)</td>
</tr>
<tr>
<td>Jan Hughes</td>
<td>George Batch (NASP)</td>
</tr>
<tr>
<td>Tim Keith</td>
<td>Kristin Hawley (Division 12)</td>
</tr>
<tr>
<td>Joel Levin</td>
<td>Steve Quintana (Cultural Diversity, Division 17)</td>
</tr>
<tr>
<td>Arthur Reynolds</td>
<td>Bruce Wampold (Division 17)</td>
</tr>
<tr>
<td>Sylvia Rosenfield</td>
<td>John Weisz (Division 12)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multicultural Issues</th>
<th>Journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colette L. Ingraham</td>
<td>Journal of Educational &amp; Psychological Consultation</td>
</tr>
<tr>
<td>Evelyn Oka</td>
<td>Emilia C. Lopez, Editor</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>School-Based Prevention Programs</th>
<th>School Psychology Quarterly</th>
</tr>
</thead>
<tbody>
<tr>
<td>George Bear</td>
<td>Rik Carl D'Amato, Editor</td>
</tr>
<tr>
<td>Joe Durlak</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>School-Based Intervention Programs for Social Behavior Problems</th>
<th>School Psychology Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shane Jimerson</td>
<td>Susan M. Sheridan, Editor</td>
</tr>
<tr>
<td>Greg Waas</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Academic Intervention Programs</th>
<th>Qualitative Methods Committee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virginia Berninger</td>
<td>Bonnie Nastasi</td>
</tr>
<tr>
<td>Ed Shapiro</td>
<td>Steve Schensul</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Family Intervention Programs</th>
<th>Comprehensive School Health Care</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cindy Carlson</td>
<td>Rune Simeonsson</td>
</tr>
<tr>
<td></td>
<td>Ronda Talley</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>School-Wide and Classroom-Based Programs</th>
<th>Research-to-Practice Committee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Craig Frisby</td>
<td>Susan Forman</td>
</tr>
<tr>
<td>Charles Maher</td>
<td>Sandra Thompson</td>
</tr>
<tr>
<td>William Strein</td>
<td></td>
</tr>
</tbody>
</table>

*Note: All members are listed in alphabetical order within domains/groups.
Table of Contents

Preface 6
  Task Force on Evidence-Based Interventions in School Psychology 6
  Future Considerations 7

Procedural and Coding Manual for Review of Evidence-Based Interventions 9

Section 1. General Considerations 9
  Procedures for Identifying Relevant Intervention Outcome Studies 9
  Literature Reviewers 9
  Task Force Domains 10
    School- and community-based intervention programs for social and behavioral problems 10
    Academic intervention programs 10
    Family and parent intervention programs 10
    School-wide and classroom-based programs 10
    Comprehensive and coordinated school health care services 10
  Literature Search Process 11
  Acceptable Sources in the Literature: Peer Review Required 11
  Coding Mixed Method Designs 11
  Include Refuting Evidence and Null Findings 11
  Obtaining Missing Information 12
  Limits of Historical Search 12
  The Literature Review Process 12
  Figure 1: Framework for the Literature Review Process 13

Section 2. Criteria for Coding Studies and Classifying Interventions and Programs 14

Section 3. Procedures for Decision-Making by the Task Force 16
  Procedures for Decision-Making by the Task Force 16
  Decision-Making Procedures 16
  Suspending the Rules 16

Section 4. Coding Manuals and Technical Coding Protocols 18

Coding Manual for Group-Based Design Intervention Research 19
  Group-Based Designs 22
    General Characteristics 22
    Key Features for Coding Studies and Rating Evidence 26
    Other Descriptive/Supplemental Criteria to Consider 41
    Coding Protocol: Group-Based Designs 45

Coding Manual for Single-Participant Intervention Research Design 63
  Single Participant Designs 66
    General Characteristics 66
PREFACE

Task Force on Evidence-Based Interventions in School Psychology

The Task Force on Evidence-Based Interventions in School Psychology (hereafter called the Task Force) was formed in 1998 and is supported by the American Psychological Association (Division 16 – School Psychology) and the Society for the Study of School Psychology (SSSP). The Task Force is also endorsed by the National Association of School Psychologists (NASP). A detailed history of the Task Force and its focus are presented in several publications [see Kratochwill & Stoiber, 2000, 2002; Stoiber & Kratochwill, 2000; and the Mini-Series in School Psychology Quarterly, 17(4)]. Current membership and contact information for the Task Force is included in Appendix A.

The Procedural and Coding Manual for Review of Evidence-Based Interventions (hereafter called the Manual) was formed to identify, review, and code studies of psychological and educational interventions for behavioral, emotional, and academic problems and disorders for school-aged children and their families. The Manual also focuses on interventions that promote health and wellness as well as prevention programs for a variety of areas. Specifically, the Manual is intended to assist in reviewing outcome research for the following purposes:

a) to identify prevention and intervention outcome studies bearing on the effects of educational and psychological prevention and intervention programs,
b) to code those studies according to Task Force criteria, while providing information on the characteristics of the interventions and the studies testing their effects,
c) to determine to what degree the interventions are evidence-based on a variety of criteria,
d) to offer the field of school psychology and related fields some guidelines for using and adopting effective programs, and
e) to provide a template for improving research in the field of psychology and education.

The primary purpose of the document is to help professionals identify, review, and code interventions that have been subjected to empirical research and evaluation. In development of the Manual we have been fortunate to take into account the concepts advanced through various drafts of the Procedural and Coding Manual for Identification of Beneficial Interventions produced by the Committee on Science and Practice by the Society for Clinical Psychology Division 12 of the American Psychological Association (Weisz & Hawley, 2002). In some sections of the Manual we have used material with permission of the Division 12 Committee. The Manual also takes into account the Criteria for Evaluating Treatment Guidelines (American Psychological Association, 2000) produced by the Template Implementation Work Group of the Board of Professional Affairs, Board of Scientific Affairs, and the Committee for the Advancement of Professional Psychology.

The Manual reflects some recent thinking and specific considerations about applying evidence-based interventions (EBIs) in school contexts as articulated elsewhere (e.g., Kratochwill & Stoiber, 2000, 2002; Stoiber & Kratochwill, 2000) and by Task Force committee members. We envision the identification of EBIs to be a long-term, ongoing process. As technologies develop and empirical evidence about an intervention accumulates, the status of the interventions in the EBI database will also evolve. The intent of the Manual is to assist the profession in developing a meaningful database on EBIs.

A major focus of the Task Force has been the development of the Manual, described in more detail below. The Manual represents an effort on the part of many individuals. We express sincere appreciation to Karen O’Connell, Task Force support person at the University of Wisconsin-Madison, for her work on numerous drafts of this document.

Similar to our Division 12 predecessors who have undertaken the task of developing a coding manual, we view this Manual as a document in progress. Knowledge and issues bearing on identification and review of EBIs are
constantly changing and the Task Force will periodically update our work. The Task Force remains open to feedback and hopes to create an ongoing forum for revision and improvement of the Manual. Feedback should be sent to Thomas R. Kratochwill at the address on the title page.

The Manual has been organized into four sections. The first section describes procedures for identifying intervention studies that are appropriate for review in the domains established by the Task Force. The second section provides specific details regarding Task Force criteria for coding intervention outcome studies and rating studies on a number of predetermined criteria. The third section describes procedures for formal decision making by the Task Force members regarding review and classification of interventions. Issues, such as resolving disagreement on coding criteria are described in this section.

The final section of the Manual includes technical codes, and is intended to provide users with a detailed understanding of how to code the evidence bearing on various interventions. This section of the Manual is the most detailed and is organized into two subcomponents (called Submanuals) that feature different methodological approaches to reviewing evidence from intervention studies. Each Submanual in this section provides an overview of the coding criteria and is accompanied by a technical coding protocol. Specifically, the following Submanuals are featured:

**Group-Based Designs.** These studies involve traditional multiple-participant investigations that classify research designs based on definitions presented by Kirk (1995). These designs include completely randomized designs and randomized block designs (both between- and within-subjects variations). The design types are applied to studies where random assignment of experimental units to intervention randomizations either is or is not included.

**Single-Participant Designs.** These studies involve traditional single-participant investigations that classify research designs based on definitions presented by Hayes, Barlow, and Nelson-Gray (1999). These designs include within-series, between-series, and combined-series and other design variations. Designs can be coded whether or not randomization is used.

**Future Considerations**

The criteria presented in the Submanuals are designed to help individuals make decisions about the quality of research evidence and empirical support of various prevention and intervention programs implemented in school and applied settings. Specifically, the Task Force is interested in making state-of-the-art conclusions pertaining to research support for various intervention programs across five domains identified later in this document. However, the Task Force also emphasizes that one of the primary purposes for identifying interventions with empirical support is so that practitioners in schools and other applied settings use these interventions. Yet, generalization from research to practice settings is not a straightforward process and often a variety of barriers exist in this process (Kazdin, Kratochwill, & VandenBos, 1986; Kratochwill & Stoiber, 2000, 2002; Stoiber & Kratochwill, 2000).

Among the variety of tactics that might be invoked to deal with the hiatus that exists between research and practice, several have been considered by the Task Force to facilitate improved generalization (Hayes et al., 1999; Kazdin et al., 1986). First, standardization of the assessment and intervention process is often associated with work on evidence-based interventions and the Task Force recommends that individuals use manuals or other procedural guidelines to facilitate the process of implementing evidence-based interventions in applied settings. Second, the Task Force recommends mechanisms for training to reduce the hiatus between research and school applied activities. One of the major goals of the Task Force will be to create a database that will be useful to training programs and professional organizations in educating practitioners, graduate students, university trainers, etc., in tactics to facilitate the dissemination of research findings to practice settings for various interventions that appear to have strong empirical support. Third, the Task Force recommends that practitioners embrace a scientist-practitioner model in their professional at work as a tactic to reduce the hiatus between research and practice. Despite the best efforts of the Task Force to address various issues surrounding intervention implementation in research, it is quite...
apparent that no series of research investigations can take into account all contextual and ecological variables that have an influence on whether an intervention is likely to be beneficial in a particular setting such as a school, classroom, etc. Therefore, it is essential that practitioners engage in self-evaluation and apply methods of empirical clinical practice to help the field understand how beneficial an EBI will be when applied in a new school context or different situation (Kratochwill & Shernoff, 2003).

The Task Force also hopes that the process of clinical replication will be invoked where practitioners can evaluate various interventions under conditions in which they are likely to be practiced (e.g., effectiveness and transportability studies conducted under real world conditions). Single-participant case study investigations, and related methods can be used to facilitate not only an evaluation of an intervention within a particular context, but also to produce information related to communication among researchers about how to revise or redesign intervention programs to make them more effective in practice settings (Hayes et al., 1999). Although a lofty goal, the Task Force hopes to promote the use of research findings in practice and during the latter stages of our activities, we hope to embrace this important training agenda (see Kratochwill & Stoiber, 2002). The Task Force formed a Research-to-Practice subcommittee in 2003 that will address this important agenda.

Finally, the Task Force wishes to emphasize the importance of taking into account multicultural and diversity issues in development of the Manual. We are requesting reviewers of the empirical literature to be especially sensitive to cultural diversity and issues surrounding the context of intervention implementation in schools and community settings. With the help of our Task Force cultural diversity domain co-chairs Colette Ingraham and Evelyn Oka and the advice of Division 17 Task Force member Steve Quintana, we trust that cultural and ethnic issues will be considered prominently within the context of intervention research.
PROCEDURAL AND CODING MANUAL FOR REVIEW OF EVIDENCE-BASED INTERVENTIONS

SECTION 1: General Considerations

Procedures for Identifying Relevant Intervention Outcome Studies

The general goal of the Task Force is to identify and review the research literature on a particular intervention. Several procedures are required to identify evidence-based interventions. First, a thorough search for relevant evidence in the research literature is required. The Task Force invites interested parties to suggest relevant literature bearing on an intervention program or technique within the domains specified in this document. Such literature might range from a single published article to a collection of studies bearing on a particular intervention proposed for Task Force consideration. Intervention programs and techniques can be implemented in schools and/or community settings. The program or technique should have relevance to school psychology and may or may not involve clinical populations. The task of identifying interventions may constitute nomination of a particular intervention or program by individuals or groups outside the Task Force. Such nominations are encouraged, with the stipulation that all available evidence bearing on effects of the nominated intervention be submitted, whether that evidence is supportive or not supportive (see below). The Task Force will also carry on its own ongoing literature review, as described in this document. Domain co-chairs will invite individuals to conduct reviews in areas that are consistent with their experience and expertise. Appendix B provides some general guidelines for conducting a literature review as well as an example literature search. Further information is available in several sources (e.g., Cooper, 1998; Fink, 1998; Hart, 2001; Lipsey & Wilson, 2001). Appendix C also provides a listing of additional resources.

Literature Reviewers

The reviewers who carry out the literature search may be members of the Task Force or others working in collaboration with the reviewers. The Task Force will focus primarily on literature review content, which is organized into five domains. Reviewers must adhere to the scope and definition of the subdomain under review and will seek assistance from the subdomain co-chairs when necessary. Initially, the Task Force developed a separate prevention domain. However, after some discussion a decision was made to identify evidence-based prevention programs within each of our five content domains.

Prevention programs focus primarily on the promotion of social, emotional, and behavioral competencies. Prevention programs reduce the incidence of mental health problems and are implemented before a problem becomes salient or is diagnosed as a disorder. Thus, all children, as well as children who are at-risk for social, emotional, and behavior dysfunction, are targeted. Although effective prevention programs have existed for some time, school-based prevention programs for issues such as substance abuse, violence, or other risk behaviors have become more prevalent in recent years (e.g., Durlak, 1997; Greenberg, Domitrovich, & Bumbarger, 2000). Programs in this domain generally include a prevention-focused curriculum that is implemented alone or as part of a more comprehensive school-wide or classroom-based program.

The decision to infuse prevention programs into each of the domains was based on the following considerations. First, the Task Force wanted an organizational framework that would allow identification of EBIs by content domain. Second, the Task Force found that prevention programs often address a continuum of intervention impact points ranging from at risk and protective factors to problems experienced by children (e.g., Minke & Bear, 2000). Third, prevention programs for one target area may be intervention programs in another domain. Thus, our decision was to include a wide range of prevention/intervention programs within our domains, and to classify them as universal, selective, or indicated programs.
Task Force Domains

The following five domains have been established by the Task Force:

**School- and community-based intervention programs for social and behavioral problems.** A wide variety of intervention programs for social/behavioral problems and disorders have been implemented in school-based settings (see Kazdin & Weisz, 2003; Morris & Kratochwill, 1998; Stoiber & Kratochwill, 1998). These intervention programs are implemented once a problem or disorder has been detected. The children that participate in these interventions usually share an essential mental health issue, circumstance, or psychological problem, such as being an abuse victim, suffering from depression, lacking anger-control, or being a child in a divorced family. In their attempt at responding to a wide range of student needs, school psychologists may implement prevention and/or intervention at the individual, group, classroom, or school-wide level.

**Academic intervention programs.** School and educational psychologists have been involved for many years in implementing intervention programs for a wide variety of academic problems (e.g., Rathvon, 1999; Shapiro, 1996). The domain will feature reviews of academic programs for the major academic areas of the curriculum (e.g., reading/literacy, mathematics, written language). Academic programs for children identified with exceptional education needs and those considered at risk within the general education curriculum will be a primary focus within this domain. Some examples of academic intervention programs include those in the area of reading (word recognition and related processes, fluency, comprehension), writing (handwriting, spelling, composition), and mathematics (computation and related processes, concepts and problem solving). Similar to the intervention programs for social and behavioral problems, consultative technologies would be appropriate for this domain.

**Family and parent intervention programs.** Family intervention programs are becoming increasingly important for school psychologists and related professionals working with children and adolescents (e.g., Christenson & Conoley, 1992; Fine & Carlson, 1992). The purposes of reviews in this area will be both family-based intervention or home-school collaboration projects that are implemented in schools or coordinated with school settings, and demonstrate a change in the child’s behavior/performance in school. Because family involvement is a significant, positive correlate of students’ academic achievement and social adjustment, schools have developed specific efforts and practices for intervening with families and for promoting home-school collaboration to enhance student success. Intervention subdomains include parent education, preschool interventions (early childhood/ family education), parent involvement, family/parent consultation, family-school partnership, and family systems.

**School-wide and classroom-based programs.** Fueled by the demands to address increasing social and behavioral problems among youth, school-wide or whole classroom programs are being advocated to provide more integrative solutions for children (Brophy, 2000). Recent literature emphasizes the critical combination of modifying teacher behavior, classroom environment, and school climate as well as actively teaching prosocial behaviors across the curriculum (e.g., Perry & Weinstein, 1998; Roeser, Eccles, & Strobel, 1998). Some examples of programs included in this domain are those addressing classroom management, system reform, inclusion, and school collaboration. Interventions such as these, which can be either preventive or remedial, typify organizational programs that would be appropriate for this domain.

**Comprehensive and coordinated school health services.** In recognition of the interrelationships among physical, psychological, social and educational functioning, school psychologists are being called upon to address the physical and mental health needs of students through participation in comprehensive and coordinated school health services (Adelman & Taylor, 1998; Power, Healthfield, McGuey, & Blum, 1999; Short & Talley, 1997). For the purposes of this review, comprehensive and coordinated school health services includes school-based or school-linked services “related to a broad spectrum of health-related problems, including chronic medical or health-related conditions (e.g., endocrine and seizure disorders, childhood cancer, asthma and allergies, diabetes, HIV/AIDS, fetal alcohol syndrome; School Psychology Review, Volumes 24 [2] & 28 [2]), psychiatric disorders (e.g., mood and behavioral disorders; Doll, 1996), and social morbidities (e.g., drug abuse, sexually-transmitted diseases, suicide,
violence; DiClemente, Hansen, & Ponton, 1996).” Authors will review research related to comprehensive health care interventions in the school health areas adopted by the Centers for Disease Control and Prevention: Health education; physical education; health services; nutrition services; healthy school environments; health promotion for staff; and parent/community involvement. To the extent these areas overlap with previously defined sections (e.g., programs for social-behavioral problems, school-wide and classroom-based), they will be treated in their respective sections. The emphasis in this domain is on primarily health-related concerns. Current models of comprehensive health care suggest the need for attention to several key factors: (a) integration of educational, health and social services; (b) interdisciplinary and inter-agency collaboration; (c) attention to ecological, contextual, and individual factors; (d) developmentally and culturally appropriate services; (e) a continuum of care ranging from prevention to treatment; (f) provision of empirically supported care; and (g) routine and systematic program evaluation (e.g., Adelman & Taylor, 1998; Dryfoos, 1995; Kolbe, Collins, & Cortese, 1997; Roberts & Hinton-Nelson, 1996). Although evaluation research is generally supportive of comprehensive health services for children and adolescents (e.g., Dryfoos, 1995; Jordan, 1996; Saxe, Cross, Lovaas, & Gardner, 1995), findings are not unambiguous. Furthermore, given the community-based nature of much of the current work, attention specifically to school-based and school-linked service provision is warranted.

Literature Search Process

Reviewers will search for relevant studies using, at a minimum, the following procedures: (a) computerized data bases (e.g., PsycInfo); (b) published reviews and meta-analyses relevant to the target domain; (c) hand searches of the most relevant journals (as determined by the reviewers) to identify appropriate studies not detected through computer search procedures; (d) studies suggested by authors of the interventions being considered; (e) studies suggested by other investigators or practitioners in the field, and (f) reference trails generated by studies thus identified. Review articles and meta-analyses serve the purpose of identifying studies to be reviewed as a starting point; the original articles on which the reviews are based must be reviewed to ensure that the results have been correctly described and appropriately interpreted with criteria invoked by the Task Force. Readers should consult Appendix B for further information on conducting literature reviews.

Acceptable Sources in the Literature: Peer Review Required

To insure some level of quality control, a consistent requirement for acceptability of the studies identified is that each must have been subjected to peer review. In most cases, this criterion will mean that the studies have been published in peer-reviewed journals; in some cases, intervention outcome studies published in books will be acceptable if there is evidence that critical peer review occurred. The Task Force will not review dissertations in their unpublished form. The peer review requirement will mean that the following will NOT be considered acceptable evidence by the Task Force: unpublished manuscripts, book chapters, and information on the Internet that has not been subjected to the peer review (except in the rare cases where critical peer review is evident).

Coding Mixed Method Designs

In some cases reviewers will be faced with the task of coding intervention research that is based on more than one methodology featured in this manual (e.g., both group and single-participant designs). In such cases, the reviewers will consult the domain co-chairs who will recommend that (a) either one primary methodology is coded, with the secondary methodology coded on relevant criteria, or (b) that both methodologies be fully coded.

Include Refuting Evidence and Null Findings

Relevant studies include not only those showing beneficial effects of the tested intervention, but also any study showing that the intervention has harmful or null effects (Kratochwill, Stoiber, & Gutkin, 2001). These characteristics will be noted when reviewers evaluate the individual studies.
Obtaining Missing Information

When data necessary for classifying interventions are not provided in the published studies (e.g., means and standard deviations), and there are no other data from which to calculate effect sizes, reviewers are provided with guidance in the Manual regarding how to calculate effect sizes. Reviewers will contact study author(s) and request the missing data when additional information is needed to code the study. Until and unless author(s) provide the necessary data, those studies will not be considered as supporting evidence for the interventions described therein.

Limits of Historical Search

Typically, all research studies in existence bearing on a prevention or an intervention domain will be included in the review. Studies dating back 25 years or more will typically not be included unless the study is of special significance.

The Literature Review Process

Generally, members of the Task Force and reviewers will follow a structured process for conducting the literature review. Figure 1 presents the general framework for this process.
Select Research Domain

Select Review Focus

Intervention Program Focus
(e.g., “self-modeling treatment”)

Problem/Issue Focus
(e.g., drug treatment/prevention/issue)

Use Best Practices in Literature Review

Select Manual Focus

Group

Single-Participant

Complete Coding Forms for Each Study

Complete Summary Coding Form

Produce Report

Tests in Schools/Practice

Practice-Research Network

Develop Practice Guidelines

Organization of Research Domains

Identification of Research

Review of Studies

Evaluation and Analysis of Research Synthesis

Summing Up: Interpretation, Presentation, and Dissemination

Figure 1. Framework for the Literature Review Process.
SECTION 2: Criteria for Coding Studies

Once outcome studies relevant to a particular intervention have been identified, the studies will undergo review and a rating will be made about the type of evidence that is available for each prevention/intervention program. Task Force reviews will follow one or more of the four manual guidelines in this document (see Section 4). In this first section of the Manual the Task Force lists criteria for reviewing the evidence in support of intended intervention effects and provides some operational definitions of various terms. More specifically, the Task Force expands the Division 12 criteria based on additional criteria presented by Crane (1998) and Troia (1999) in their literature reviews and Lipsey and Wilson (2001) in their discussion of meta-analysis.

Because of potential problems in relying on experimental design features to determine whether an intervention produced the intended effects on the participants who are presumed to benefit (i.e., children, families, teachers, etc.), the method the Task Force has chosen for evaluation is delineated into three types: (a) general characteristics, (b) key features, and (c) supplemental descriptive information. The first type of coding consideration in the Manual addresses the general methodological qualities and statistical procedures used in determining the effects of an intervention. These criteria refer to the credibility of the intervention approach and the quality of the outcome evaluation. Reviewers will take into account the overall methodological quality of the outcome evaluation, and in particular, the use of appropriate evaluation methods that correspond to features of the school environment. This first set of information should be helpful in providing a context for understanding the research conducted to demonstrate intervention effects (for example, whether the study incorporated a group or single-participant design is noted).

Nine key features of an intervention study will be rated on a 4-point scale to specify level of evidence (i.e., 3 = strong evidence/support, 2 = promising evidence/support, 1 = marginal or weak evidence/support, 0 = no evidence/support). These nine key features focus both on internal and external validity criteria, as well as on features considered important for school- or field-based implementation. The nine key features include: (a) Measurement, (b) Comparison Group, (c) Primary/Secondary Outcomes Significant, (d) Educational/Clinical Significance, (e) Durability of Effects, (f) Identifiable Intervention Components, (g) Implementation Fidelity, (h) Replication, and (i) School- or Field-based Site (see American Institutes for Research & Herman, 1999, An Educators’ Guide to Schoolwide Reform for an example of this type of rating).

Generally, a premium is placed on studies incorporating a comparison group that permits judgment of the merits of an intervention. Because experience with school-based or field-based investigations has often revealed practical problems in implementing random assignment (Lipsey & Cordray, 2000), consideration will be given to alternative methodological and statistical strategies incorporated to contend with internal validity threats. In addition, reviewers will take into account various internal validity issues that might influence interpretation of outcomes even when random assignment has been used (e.g., resentful demoralization, differential attrition) – see Shadish, Cook, and Campbell (2002) and Levin and O’Donnell (1999).

Because descriptive information about intervention implementation and patterns of change often cannot be readily incorporated into standard statistical analysis or quantitative ratings, another type of information that will be made available to readers describes other important qualities of the study that do not lend themselves well to a rating scale structure. This type of information, however, may be particularly useful for individuals who are interested in evaluating whether the intervention is matched well to specific program needs in applied settings. Included within this third set of descriptions is information on external validity. Hence, the third set of codes should provide supplemental considerations for the reader, who can judge the beneficial or evidence-based dimensions of a
particular intervention or program in terms of specific needs, qualities, or expectations. For this type of information, descriptions related to qualities of the study are provided rather than a judgment or rating of evidence.
SECTION 3: Procedures for Decision-Making by the Task Force

Procedures for Decision-Making by the Task Force

Given the virtually infinite array of forms that intervention outcome research may take, and the great diversity of possible findings across various intervention conditions and outcome measures, it is difficult to produce a set of coding, decision, and classification rules that will address all possible studies and outcomes. Accordingly, application of the manualized procedures in this document needs to be complemented by Task Force committee deliberations on the state of the evidence bearing on each intervention under consideration. Thus, intervention review and classification will proceed according to the following procedures.

Decision-Making Procedures

For each intervention under consideration, a “subcommittee” of the Task Force (a group with experience and expertise bearing on that intervention domain, as noted above and typically consisting of the Task Force subdomain co-chairs and with the consultation of the conceptual/methodological domain co-chairs) will apply the coding and classification criteria according to the Manual. The subcommittee will report to the Task Force chair, describing how the intervention fared according to manual requirements, and noting any unusual features, important issues, concerns, or problems that arose during the review. All this information will be discussed (orally, via e-mail, or through some equivalent process) between the subdomain co-chairs, and a formal vote may be taken on how the intervention should be classified along evidence-based criteria by the Task Force chair if the subdomain co-chairs with the advise of the conceptual and methodological domain co-chairs cannot reach consensus. In this vote, a simple majority of a quorum (i.e., 75% of the voters at the time) will determine the classification decision.

Under this procedure, reviewers apply the review and coding procedures specified in the Manual, but that process is complemented by formal committee review to discuss unusual aspects, special concerns, or issues not covered in the manual, and to reach a final classification decision. To insure openness and invite public scientific scrutiny, the proceedings of the meeting, a summary of the discussion, and the result of the vote, will be made (e.g., in periodic “update” articles, or presented via the Task Force web site). The proceedings available will include a summary of the issues and topics discussed as well as results of votes, but names will not be linked with particular statements or opinions.

Suspending the Rules

Although there are clear advantages to applying explicit rules for coding studies and categorizing interventions, no set of rules is likely to fit all circumstances equally well. Accordingly, there are likely to be cases in which the Task Force needs to consider suspending the rules. First, there may be instances in which a particular prevention/intervention program does not meet many criteria for EBIs, but in which extenuating circumstances make it unreasonable to apply all the criteria rigidly. As one example, consider the subset of intervention outcome studies conducted in the 1970s, which did not include tests of adherence to intervention protocols; given that such tests were not a standard part of outcome research at the time, our requirement that there be adherence tests may be suspended for these early studies. As another example, there may be good reason to suspend rules about sample size for interventions that address very rare conditions (e.g., autism, selective mutism). These, and a variety of other decisions regarding application of Task Force rules, will be resolved by committee vote and such decisions will be noted for the particular study under consideration. It is also possible that there may be intervention programs for which the supporting research does meet all the criteria specified in the Submanuals, but for which Task Force members have remaining concerns that raise doubts about whether the intervention is, in fact, as good as it
appears (e.g., where all outcome measures were completed by the therapist/consultant who is also the intervention-developer). In this situation, too, the views and concerns of subcommittee members (including any disagreements among members) will be presented to the Task Force Chair for discussion and possibly, a vote by the full membership on whether to suspend a specific rule in a specific instance. Decisions will be determined by the “majority of a quorum” rule noted above.
SECTION 4: Coding Manuals and Technical Coding Protocols
CODING MANUAL FOR

GROUP-BASED INTERVENTION RESEARCH DESIGN

Task Force on Evidence-Based Interventions in

School Psychology
The Coding Manual for Group-Based Intervention Research Design was created by the Task Force on Evidence-Based Interventions in School Psychology.

The following individuals contributed to development of the Coding Manual:

* Sandra Christenson, PhD
  University of Minnesota

* Joseph A. Durlak, PhD
  Loyola University – Chicago

* Thomas R. Kratochwill, PhD
  University of Wisconsin-Madison

* Joel R. Levin, PhD
  University of Arizona

* Gretchen Lewis-Snyder •
  University of Wisconsin-Milwaukee

* Karen Callan Stoiber, PhD
  University of Wisconsin-Milwaukee

* Elisa Steele Shernoff •
  University of Wisconsin-Madison

* Ronda Talley, PhD, MPH
  Tri-T Associates, Inc.

* Greg A. Waas, PhD
  Northern Illinois University

Stacey Bates •
University of Texas

Laura Gul •
University of Texas

Dan Bolt, PhD
University of Wisconsin-Madison

Bonnie Nastasi, PhD
Institute for Community Research, CT

Cindy Carlson, PhD
University of Texas

Arthur Reynolds, PhD
University of Wisconsin-Madison

Diane Cox •
University of Texas

Steve Schensul, PhD
University of Connecticut Health Center

Maria Fischel •
University of Texas

Carmen Valdez •
University of Texas

The Task Force also acknowledges contribution of the Interdisciplinary Qualitative Research Committee:

<table>
<thead>
<tr>
<th>Name</th>
<th>Specialty Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bonnie Nastasi, Co-Chair</td>
<td>School Psychology</td>
</tr>
<tr>
<td>Stephen Schensul, Co-Chair</td>
<td>Anthropology</td>
</tr>
<tr>
<td>Doug Campbell</td>
<td>Educational Research</td>
</tr>
<tr>
<td>Denise DeZolt</td>
<td>School Psychology</td>
</tr>
<tr>
<td>David Fettermann</td>
<td>Anthropology</td>
</tr>
<tr>
<td>Karen Harris</td>
<td>Special Education/Educational Psychology</td>
</tr>
<tr>
<td>Evelyn Jacob</td>
<td>Anthropology</td>
</tr>
<tr>
<td>Colette Ingraham</td>
<td>School Psychology</td>
</tr>
<tr>
<td>Margaret LeCompte</td>
<td>Sociology</td>
</tr>
<tr>
<td>Joel Meyers</td>
<td>School Psychology</td>
</tr>
<tr>
<td>Evelyn Oka</td>
<td>School Psychology</td>
</tr>
<tr>
<td>Mittie Quinn</td>
<td>School Psychology</td>
</tr>
<tr>
<td>Jean J. Schensul</td>
<td>Anthropology</td>
</tr>
<tr>
<td>Cherie Tyler</td>
<td>School Psychology</td>
</tr>
<tr>
<td>Stephen Truscott</td>
<td>School Psychology</td>
</tr>
</tbody>
</table>
* Designates membership on the original Manual Subcommittee.
• Designates graduate student

For further information please contact:

Thomas R. Kratochwill, PhD
Chair, Task Force on Evidence-Based Interventions in School Psychology
School Psychology Program
1025 W. Johnson Street
University of Wisconsin-Madison
Madison, WI 53706-1796
tomkat@education.wisc.edu
(608) 262-5912
GROUP- BASED DESIGN

I. General Characteristics

A. General Design Characteristics

Traditional multiple-participant or group-based studies will be classified according to the type of research design, based on Kirk's (1995) terminology. If authors provide sufficient information, the categories will be further subdivided as follows:

A1. Random-assignment designs. To receive a code for random-assignment design, authors must specify that random assignment to treatment conditions occurred. If such a statement is not made, then nonrandom assignment is assumed.

A1.1 Completely randomized design, in which at least one type of comparison group is used (e.g., nonintervention, attention control, alternate intervention). Participants are randomly assigned to intervention and comparison conditions.

A1.2 Randomized block design (between-participants variation), in which at least one type of comparison group is used. Participants are initially blocked, or matched, on the basis of some pre-study variable(s) and then randomly assigned to intervention and comparison conditions. For example, a study examining the effects of a classroom level intervention might use this type of design because it is difficult to randomly assign students to a classroom, or classrooms to a school. Instead, the investigators can “match” classrooms based on characteristics such as teacher to student ratio, student reading achievement, student attendance rates, and number of students displaying frequent disruptive behavior. Once two “similar” classrooms are found, or matched, they can be randomly assigned to an intervention or comparison group. This matching process continues until the investigators reach the desired sample size.

A1.3 Randomized block design (within-participants variation), in which at least one type of comparison group is used. Participants are administered both intervention and comparison conditions in different random orders, in either a single-factor layout (i.e., intervention vs. comparison conditions only) or in the context of a multiple-factor design (where the other factors consist of either researcher-manipulated or participant factors). If one or more between-participant factors are combined with one or more within-participant factors, the design is technically referred to as a “split-plot” design, but for present purposes the A1.3 category is sufficient.

A1.4 Randomized hierarchical design. In these designs, treatments are nested instead of being crossed.

A2. Nonrandom-assignment designs. These designs are identical to the random assignment designs described above, except there is no random assignment.

A2.1 Nonrandomized design, in which at least one type of comparison group is used. Participants are nonrandomly assigned to intervention and comparison conditions.

A2.2 Nonrandomized block design (between-participants variation), in which at least one type of comparison group is used. Participants are initially blocked, or matched, on the basis of some pre-study variable and then nonrandomly assigned to intervention and control conditions.
A2.3  **Nonrandomized block design** (within-participants variation), in which at least one type of comparison group is used. Participants are administered both intervention and comparison conditions in different nonrandom orders, in either a single-factor layout (i.e., intervention vs. comparison conditions only) or in the context of a multiple-factor design (where the other factors consist of either researcher-manipulated or participant factors). If one or more between-participant factors are combined with one or more within-participant factors, the design is technically referred to as a "split-plot" design, but for present purposes the A2.3 category is sufficient.

A2.4  **Nonrandomized hierarchical designs.** This design is the same as the hierarchical design noted above but random assigned is not used.

A2.5  **Optional coding of Quasi-experimental designs** (see Appendix D)

A3.  **Overall confidence of judgment on how participants were assigned.**

A3.1  **Very low** (little basis)
A3.2  **Low** (guess)
A3.3  **Moderate** (weak inference)
A3.4  **High** (strong inference)
A3.5  **Very high** (explicitly stated)
A3.6  **N/A**
A3.7  **Unknown/unable to code**

B. **Statistical Treatment** (B1 through B6 are coded for each statistical test that is used in a study).

B1.  **Appropriate unit of analysis.** The unit of analysis corresponded to the unit of intervention (randomization and treatment administration) so that the assumption of independence of errors was not violated and Type I error probability was adequately controlled. For example, if a school-wide violence prevention program was being investigated, then whole schools would be the units of analysis—whole schools would be considered the “participant,” and the outcome data would be collected and analyzed at the whole school level or by a hierarchical model that separates school and student level effects. By contrast, individual therapy is administered to individual children; therefore, the unit of analysis should be the child (e.g., the child is the “participant”).

B2.  **Familywise error rate controlled.** There are many ways to control Type I error. However, the purpose of this code is to determine whether some procedure was used to control the familywise or experimentwise Type I error probability when multiple statistical comparisons were made (e.g., use of a MANOVA procedure when multiple related outcomes are evaluated, use of a controlled multiple-comparison procedure used when more than one follow-up comparison is made).

B3.  **Sufficiently large N.** The number of participants (after attrition) in each condition is sufficient to yield enough statistical power for detecting effects of interest or importance. Table 1 provides a guide for the N required for a small, medium, and large effect size, given the alpha level. When using Table 1, select a statistical test and an alpha level for the analyses in question. This table reflects that larger Ns are needed to detect smaller effects.

B4.  **Total size of sample** (start of the study).

B5.  **Intervention group sample size.**

B6.  **Control group sample size.**
Table 1

N for Small, Medium and Large ES at Power = .80 for _ = .01, .05, and .10

<table>
<thead>
<tr>
<th>Test</th>
<th>.01</th>
<th>.05</th>
<th>.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean diff</td>
<td>586</td>
<td>95</td>
<td>38</td>
</tr>
<tr>
<td>Sig r</td>
<td>1165</td>
<td>127</td>
<td>44</td>
</tr>
<tr>
<td>r dif</td>
<td>1165</td>
<td>127</td>
<td>44</td>
</tr>
<tr>
<td>P = .5</td>
<td>1165</td>
<td>127</td>
<td>44</td>
</tr>
<tr>
<td>1df</td>
<td>1168</td>
<td>130</td>
<td>38</td>
</tr>
<tr>
<td>2df</td>
<td>1388</td>
<td>154</td>
<td>56</td>
</tr>
<tr>
<td>3df</td>
<td>1546</td>
<td>172</td>
<td>62</td>
</tr>
<tr>
<td>4df</td>
<td>1675</td>
<td>186</td>
<td>67</td>
</tr>
<tr>
<td>5df</td>
<td>1787</td>
<td>199</td>
<td>71</td>
</tr>
<tr>
<td>6df</td>
<td>1887</td>
<td>210</td>
<td>75</td>
</tr>
<tr>
<td>ANOVA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2g</td>
<td>586</td>
<td>95</td>
<td>38</td>
</tr>
<tr>
<td>3g</td>
<td>464</td>
<td>76</td>
<td>30</td>
</tr>
<tr>
<td>4g</td>
<td>388</td>
<td>63</td>
<td>25</td>
</tr>
<tr>
<td>5g</td>
<td>336</td>
<td>55</td>
<td>22</td>
</tr>
<tr>
<td>6g</td>
<td>299</td>
<td>49</td>
<td>20</td>
</tr>
<tr>
<td>7g</td>
<td>271</td>
<td>44</td>
<td>18</td>
</tr>
<tr>
<td>Multi R</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2k</td>
<td>698</td>
<td>97</td>
<td>45</td>
</tr>
<tr>
<td>3k</td>
<td>780</td>
<td>108</td>
<td>50</td>
</tr>
<tr>
<td>4k</td>
<td>841</td>
<td>118</td>
<td>55</td>
</tr>
<tr>
<td>5k</td>
<td>901</td>
<td>126</td>
<td>59</td>
</tr>
<tr>
<td>6k</td>
<td>953</td>
<td>134</td>
<td>63</td>
</tr>
<tr>
<td>7k</td>
<td>998</td>
<td>141</td>
<td>66</td>
</tr>
<tr>
<td>8k</td>
<td>1039</td>
<td>147</td>
<td>69</td>
</tr>
</tbody>
</table>

Note. ES = population effect size, Sm = small, Med = medium, Lg = large, dif = difference, ANOVA = analysis of variance. a

For studies using qualitative research methods, code B7 and B8

B7. Coding. The use of a systematic and clearly articulated approach to coding of qualitative data that can involve inductive and/or deductive analysis, and meets the following criteria:

B7.1 Coding scheme is linked to study’s theoretical-empirical basis
B7.2 Procedures for ensuring consistency of coding are used (e.g., dialogue among coders to ensure consistency of meaning and clarify discrepancies; testing the consistency of code application across coders and data).
B7.3 Progression from abstract concepts to empirical exemplars is clearly articulated.

B8. Interactive process. The processes of data collection and analysis interact such that preliminary analytical results are subjected to further and more focused data collection for the purposes of confirmation, explanation, and validation. The processes of data collection and analysis continues until consistency and depth of understanding are satisfactorily established.
C. Type of Program

The prevention or intervention program should be classified according to the Institute of Medicine’s (1994) classification system. Universal prevention programs are designed to forestall the development of mental health problems and to promote competence. Often, in universal programs, it is theorized that everyone would benefit from the program. Selective interventions target individuals or subgroups of a population whose risk of developing a problem is significantly higher than average, or individuals who do not respond to universal prevention programs. Indicated interventions target minimal or detectable signs that foreshadow the development of a problem (IOM, 1994; Shinn, Walker & Stoner, 2002). Intervention or treatments, on the other hand, are provided immediately to treat a serious disorder or problem (Institute of Medicine, 1994).

C1. Universal prevention program
C2. Selective prevention program
C3. Targeted prevention program
C4. Intervention/Treatment
C5. Unknown

D. Stage of Program

Prevention or intervention programs will also be classified according to the developmental stage of the program. Rossi and Freeman (1995) classify programs as model/demonstration, early stage, or established/institutionalized programs. Model/demonstration programs are those programs being implemented for the first time, and are usually small in scale and evaluated using the most rigorous methods possible. Fidelity is a major research question at this stage. Evaluations of intervention and prevention programs occurring in the early stages are likely to use mixed methods (e.g., a combination of qualitative and quantitative criteria). Finally, evaluations of established programs are less likely to investigate fidelity of implementation and are more likely to use nonrandomized designs and alternative-intervention comparison groups.

D1. Model/demonstration programs
D2. Early stage programs
D3. Established/institutionalized programs
D4. Unknown

E. Concurrent or Historical Intervention Exposure

E1. Current Exposure. Participant(s) are exposed to another intervention currently (Specify if information is available).
E2. Prior Exposure. Participant(s) were previously exposed to other interventions (Specify if information is available).
E3. Unknown. No information is available regarding concurrent or historical intervention exposure.
II. Key Features for Coding Studies and Rating Level of Evidence/Support

The second type of coding includes eight key features of an intervention study that are rated on a four-point scale to specify the level of evidence (i.e., 3=strong evidence/support, 2= promising evidence/support, 1= marginal or weak evidence/support, 0= no evidence/support).

A. Measurement

Studies will be evaluated with regard to qualities of the measures used to establish the effects of an intervention. In cases where measures are not well known or not described in enough detail to be evaluated, the coder should contact the author(s) for further information.

To receive a rating of 3 for strong evidence, studies must use measures that produce reliable scores for the majority of primary outcomes under investigation, and for the current population under study. Wallen and Fraenkel (2001) reported for research purposes, reliability should be at least .70, and preferably higher. Since .70 is the rule of thumb for a minimum, a rating of 3 for strong evidence would require that, instruments produce a reliability coefficient of .85 or higher. The reliability information should be provided including the type of reliability statistic used. In addition, to receive a rating of 3, data should have been collected using multiple methods, and collected from multiple sources, when appropriate. Finally, the investigators must have presented a case for the measures used to assess primary outcomes. In the event that multiple primary outcome measures are used, the above criteria must be met for all primary outcome measures.

To receive a rating of 2 for promising evidence, studies must use measures that produce reliable scores for the primary outcomes under investigation, and for the current population under study (i.e., reliability coefficient of at least .70). In addition, data should have been collected either (1) using multiple methods and/or (2) from multiple sources, when appropriate. A case for validity does not need to be presented. In the event that multiple primary outcome measures are used, the above criteria must be met for at least 75% of the primary outcome measures.

To receive a rating of 1 for weak evidence, studies should use measures that produce somewhat reliable scores for the primary outcomes under investigation, and for the current population under study (i.e., reliability coefficient of at least .50). In addition, data may have been collected either (1) using multiple methods and/or (2) from multiple sources; however, this is not required for a rating of 1. A case for validity does not need to be presented. In the event that multiple primary outcome measures were used, the above criteria must be met for at least 50% of the primary outcome measures.

A rating of 0 indicates that the measures did not produce reliable scores or produced scores with low reliability (<.50), AND/OR data were not collected using multiple methods, AND/OR data were not collected from multiple sources. For example, if the outcome measure is academic achievement and the student's reported achievement is used as the only outcome measure, this would be scored as a 0 (because it is not reliable and only one source was used).

A1. Use of outcome measures that produce reliable scores. Outcome measures used were supported by evidence of such psychometric characteristics as inter-rater reliability, test-retest reliability, internal consistency, for the population under study. Therefore, when possible, reliability estimates should be reported for the measures used with the samples in the study. This information must either be reported or referenced in the article to receive this code. See Table 2 for information on ways in which reliability can be assessed. Sometimes, investigators combine several measures by using factor analytic procedures to create a single score for a particular construct. This procedure is generally considered a more robust procedure, although the internal consistency of the composite score may not be as strong. In these cases, coders should consider the available reliability and validity evidence and use discretion in their judgment. For example, if researchers combine three
scores (≥ .85) using confirmatory factor analysis to create a composite score, such that all three scores where derived using multiple methods and from multiple sources, and such that all three scores loaded significantly on one factor, then coders could consider this strong evidence of sound measurement procedures.

Caveat on norm-referenced measures. Often times, researchers use published, norm-referenced measures to assess change. While these instruments are generally considered “reliable,” it is important to note that some norm-referenced assessments may not produce reliable scores for certain populations. For example, some studies may use well-known, norm-referenced measures on populations not well represented in the norming process. Even if the population in question was represented in the norming process, it does not necessarily mean the instrument produced reliable scores for that group. In addition, researchers will sometimes use less reliable subscales of published norm-referenced instruments as key outcome measures. Ideally, researchers should have calculated a reliability coefficient for all instruments used to measure change to ensure that the measure produced reliable scores for the population under study, even when well-known and widely used measures were used. When authors do not provide these data, coders would consider the absence of that criterion when making their reliability rating.

A2. Multi-method. Multiple (i.e., at least two) assessment methods or approaches were used (e.g., observational, self-reports, teacher ratings). This rating is not always applicable. For example, directly observable incidence rates may not require a multi-method assessment.

A3. Multi-source. Measures were obtained from multiple (i.e., at least two) sources (e.g., teachers, parents, self). This rating is not always applicable. Directly observable incidence rates would not necessarily require a multi-source assessment.

A4. Validity of measures reported. Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of tests (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Validation involves gathering evidence to provide a sound scientific basis for the interpretation of scores. It is important that authors note the way in which validity of their primary outcome measures was assessed. Table 2 provides information on different types of validity that can be assessed, as well as some important cultural considerations in assessing validity. Observable incidence and/or occurrence rates, such as school attendance rates and homework completion rates, and well-known standardized, norm-referenced assessments will be considered valid measures. However, coders should consider the cultural considerations listed in Table 2 and use discretion in coding measure validity.
### Table 2

*Methods of Checking Validity and Reliability*

<table>
<thead>
<tr>
<th>Validity Type</th>
<th>Cultural Considerations</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evidence based on test content (Content Validity)</td>
<td>Differences in language and meaning</td>
<td>Expert judgment</td>
</tr>
<tr>
<td></td>
<td>Relevance of items for different cultural groups</td>
<td>Translation and back translation procedures</td>
</tr>
<tr>
<td></td>
<td>Culturally defined response styles</td>
<td></td>
</tr>
<tr>
<td>Evidence based on response processes</td>
<td>Individuals are responding to the measure as represented by the construct</td>
<td>Conduct an analysis of individual items</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Question responders about the meaning of their responses</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Document related aspects of performance as related to the construct being measured</td>
</tr>
<tr>
<td></td>
<td>Analyze the relationship among parts of the test and between the test and other constructs/variables</td>
<td>Assess how observers/judges record and evaluate data related to the response processes</td>
</tr>
<tr>
<td>Evidence based on relations to other variables (Criterion Validity)</td>
<td>Accuracy of predictions across different cultural groups</td>
<td>Relate to another measure of the same variable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Relationship between tests scores and relevant criteria are modeled by regression lines and compared across groups.</td>
</tr>
<tr>
<td>Evidence based on internal structure (Construct Validity)</td>
<td>Construct defined in a culturally relevant fashion</td>
<td>Assess evidence on predictions made from theory</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Factor analysis conducted using data collected from population under investigation.</td>
</tr>
<tr>
<td>Evidence based on consequences of testing</td>
<td>Individual benefits will be derived from the intended use of scores (e.g., beneficial interventions)</td>
<td>Assess evidence of differential benefits of measures for specific groups</td>
</tr>
<tr>
<td>Type</td>
<td>Content</td>
<td>Time Interval</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------</td>
<td>---------------</td>
</tr>
<tr>
<td>Test-retest</td>
<td>Identical</td>
<td>Varies</td>
</tr>
<tr>
<td>Equivalent forms</td>
<td>Different</td>
<td>None</td>
</tr>
<tr>
<td>Internal Consistency</td>
<td>Different</td>
<td>None</td>
</tr>
<tr>
<td>Observer agreement</td>
<td>Identical</td>
<td>None</td>
</tr>
</tbody>
</table>


### B. Comparison Group

In judging the merit of an intervention, there must be a comparison group of some type against which to compare the intervention. The intervention in question could have been shown to be superior to a no intervention or wait list control, an attention or placebo control, or an alternate intervention. If the intervention in question was compared to an alternate intervention and to a control condition in one study, it may be shown superior to the control condition but not the alternate intervention. Evidence ratings will consider all available information related to the comparison group findings.

To be rated a **3 for strong evidence**, at least one type of “active” comparison group (e.g., typical intervention, attention placebo, intervention element placebo, alternative intervention, pharmacotherapy) must be used. Initial group equivalency must be established, preferably through random assignment of participants to intervention conditions. There must be evidence that change agents were counterbalanced, as well as the study must meet the criteria for equivalent mortality and low attrition at post, and if applicable, at follow-up.

To receive a rating of **2 for promising evidence**, at least a "no intervention group" type of comparison must have been used (e.g., no intervention, wait list/delayed intervention, minimal contact). In addition, there must be evidence for at least two of the following: (1) counterbalancing of change agents, (2) group equivalence established, or (3) equivalent mortality with low attrition. If equivalent mortality is not demonstrated, an intent-to-intervene analysis must have been conducted, resulting in a finding of no significant group differences between the control and treatment groups.

A rating of **1 for weak evidence** would require a comparison group and at least one of the following: (1) counterbalancing of change agents, (2) group equivalence established, or (3) equivalent mortality with low attrition. If equivalent mortality is not demonstrated, an intent-to-intervene analysis must have been conducted,
resulting in a finding of no significant group differences between the control and treatment groups.

A rating of 1 would require a comparison group (e.g., a waiting list); however, no group equivalence procedures were used (according to the methods described next).

A rating of 0 indicates that no efforts were made to ensure group equivalence.

**B1. Type of comparison group will be classified as:**

- **B1.1 Typical intervention:** control receives a typical intervention in a school setting without special additions that constitutes the intervention of interest
- **B1.2 Typical intervention (other):** Specify
- **B1.3 Attention placebo:** control receives attention, discussion, or a weaker form of the intervention.
- **B1.4 Intervention element placebo:** control receives target intervention minus the active ingredient linked to therapeutic change.
- **B1.5 Alternative intervention:** Control receives another intervention (other than typical intervention) or at a lower dosage than the established intervention (e.g., planned variation studies). This is applicable when the alternative intervention is not expected to be as effective as target intervention.
- **B1.6 Pharmacotherapy.**
- **B1.7 No intervention:** no evidence of treatment or attention
- **B1.8 Wait list/delayed intervention:** contact limited to application, screening, pre or posttest.
- **B1.9 Minimal contact:** includes participation in intake interview, given instructions, but not wait listed.
- **B1.10 Unable to identify type of comparison group**

**B2. Overall confidence rating on judgment of type of comparison group**

- **B2.1 Very low** (little basis)
- **B2.2 Low** (guess)
- **B2.3 Moderate** (weak inference)
- **B2.4 High** (strong inference)
- **B2.5 Very high** (explicitly stated)
- **B2.6 Unable to identify comparison group**

**B3. Counterbalancing of change agents.** All participants who received intervention from a single therapist/consultant/mediator/instructor or multiple agents were counterbalanced across intervention conditions to avoid agent-by-condition confound. This coding means, for example, that any effects an individual therapist/consultant/teacher/mediator may have produced in the intervention condition also affected participants in the comparison condition. A statistical analysis that controls for “change agent” or “classroom” effect is acceptable.

**B4. Group Equivalence Established**

- **B4.1 Random assignment.** To be coded for random assignment, participants must have been assigned randomly to intervention and control groups in group-based designs. Assignment may have been completely random, or random with blocking on, for example, level of treated problem, age, social economic level, pretest measures and/or IQ scores, number of suspensions, or other variables deemed important by the study authors. Threats to internal validity that random assignment does not rule out are considered in determining the level of evidence to support random assignment.
B4.2 Posthoc matched set. Participants were paired and matched on a given set of characteristics such as age, gender, ethnicity, ability, etc., and then assigned randomly to conditions.

B4.3 Statistical matching. This procedure is accomplished by adjusting participants’ scores by one or more covariates.

B4.4 Posthoc test for group equivalence. Statistical tests were conducted to demonstrate that, prior to the intervention, the intervention and control group(s) were equivalent. One must be cautious to have enough power with this procedure.

B5. Equivalent mortality with low attrition (less than 20% for post; 30% for follow-up). The number of participants in each treatment group at the conclusion of the study was approximately the same (i.e., less than 20% attrition at post assessment and less than 30% at follow-up assessment). Nonequivalent mortality rates may be associated with negative treatment characteristics or important differences between those participants who remain in the study and those who withdraw.

Levels of attrition will be reported for all groups/participants in the study, to the extent that the study authors provide this information. Reviewers will also note how attrition was accounted for in the data analysis (e.g., initial versus terminal intervention sample as well as nonintervention dropouts, analysis intent-to-intervene participants or completers only). Reviewers will report on whether intent-to-intervene analyses were carried out, and if so, what the findings were and how they differed from findings of “as-intervened” analyses.

C. Primary/Secondary Outcomes Are Statistically Significant

One important criterion is whether the intervention demonstrates a statistically significant effect on the outcome measures in the intervention group or condition. The Task Force will consider the guidelines presented by Wilkinson and the Task Force on Statistical Inference (1999) of the APA Board of Scientific Affairs. Although statistical significance is a necessary criterion for intervention outcome efficacy, it is not sufficient in that statistically significant effects do not guarantee effects of practical importance. The statistical significance, the effect sizes, and whenever possible, the statistical power, from all primary outcome measures will be assessed in judging the evidence to support a primary outcome. Outcomes will be considered statistically significant if they reach an alpha level of .05 or less. Familywise and experimentwise error rates will be considered.

To be considered a “primary outcome,” the outcome measure should represent a valid and appropriate indicator for the type of prevention or intervention. In other words, “primary outcomes” are those that reflect the ultimate goal(s) of the intervention. For example, for interventions that target a behavior-linked problem such as drug use or sexual activity, at least one outcome measure must assess behavioral functioning or adjustment status; an outcome restricted to knowledge assessment would not be sufficient. Similarly, if an intervention targets reading improvement, than the outcome must measure reading performance. For both of these intervention types, outcomes restricted to attitudes would not be sufficient to meet the criteria of a primary outcome. Outcomes significant at the $p < .05$ level will be described in the “Primary/Secondary Outcomes Significant” table. Reviewers will note any nonsignificant and/or negative outcomes associated with the intervention in a separate table (see “Null Findings/Negative Outcomes”).

To receive a rating of 3, strong evidence, an appropriate statistical analysis must have been conducted, including appropriate units of analysis, familywise/experimentwise error rate controlled (if applicable), and a sufficiently large $N$. A study must show significant primary outcomes for at least 75% of the total primary outcome measures for each key construct. Please note that “75%” refers to a single primary outcome construct, not a sum of the total primary outcome measures. For example if an increase in reading
performance and reduced disruptive behavior were identified as primary outcomes, and there were 2 measures of reading and 4 measures of disruptive behavior, then both measures of reading and 3 of 4 measures of disruptive behavior must demonstrate statistical significance in the desired direction to receive a rating of 3, strong evidence. Measured outcomes must also reflect a moderate effect size (See Tables 3-6). In addition, if primary outcomes are dependent upon related outcomes as suggested by the underlying theory, statistical significance must be demonstrated for these outcomes. If a statistical procedure indicating the magnitude of an intervention effect was not calculated, the coder should attempt to compute it based on available data.

To receive a rating of 2, promising evidence, an appropriate statistical analysis must have been conducted, including appropriate units of analysis, familywise/experimentalwise error rate controlled (if applicable), and a sufficiently large $N$. A study must show significant primary outcomes for at least 50% to 74% of the total primary outcome measures for each key construct (e.g., at least 2 out of 4 primary outcome measures for a given construct show statistically significant change in the desired direction). In addition, if primary outcomes are dependent upon related outcomes as suggested by the underlying theory, statistical significance must be demonstrated for these outcomes.

To receive a rating of 1, weak evidence, an appropriate statistical analysis must have been conducted, including appropriate unit of analysis, familywise/experimentalwise error rate controlled (if applicable). A study must show significant primary outcomes for between 25% and 49% of the total primary outcome measures for any key construct (i.e., at least 1 out of 4 outcome measures show statistically significant change in the desired direction for a given primary outcome construct).

A rating of 0 would indicate that none of the above criteria were met. Significant primary outcomes will be coded according to the following categories:

**C1. Evidence of appropriate statistical analysis for primary outcomes.** Coders will consider whether an appropriate statistical analysis was conducted using the information gathered in Statistical Treatment (Section I, Criterion B). In addition, coders will consider any glaring errors and oversights in the statistical analysis that may not have been captured in the Statistical Treatment section (e.g., an obvious violation of an assumption, such as nonindependence of error components in an Analysis of Variance).

C1.1 Appropriate unit of analysis
C1.2 Familywise/experimentalwise error rate controlled
C1.3 Sufficiently large $N$

**C2. Percentage of primary outcomes that are statistically significant**

C2.1 Outcome. Specify the outcome being tested, noting primary outcomes before secondary outcomes. All outcomes are listed in alphabetical order.

C2.2 Primary vs. Secondary. Specify whether the outcome was primary or secondary.

C2.3 For whom the outcome(s) were significant
Child
Teacher
Parent/significant adult
Ecology (e.g., classroom context/peers).

C2.4 What changed
Behavior: Specify
Attitude: Specify
Knowledge: Specify  
Dynamics: Specify  
Other: Specify  

C2.5 Source/type of measurement  
Self report: Specify  
Parent report: Specify  
Teacher report: Specify  
Peer ratings/nominations: Specify  
Observation: Specify  
Test: Specify  
Other: Specify  
Unknown/information not provided  

C2.6 Treatment Information. Specify which treatment groups are being compared.  

C2.7 Outcome measure used and reliability. Specify which measure was used to investigate the outcome, and list the reliability. If subscales were analyzed individually, then treat them as individual measures of a given construct. If measures or scales were combined, then treat the combined score as a single measure of a given construct, and list the reliability accordingly.  

C2.8 Effect Size. The effect size will be recorded as provided by authors. If the authors do not provide the effect size, then the effect size will be calculated by the reviewer(s). (See Tables 3-6 for instructions on how to calculate selected effect sizes and magnitude guidelines.) In addition, the type of data the effect size is based on will be reported, as well as the strength of the statistic (e.g., small, medium and large). Appendix D also provides general resources that will assist reviewers in calculating and interpreting effect sizes.  

C2.9 Power. Statistical power considerations should be taken into account.  

C3. Evidence of appropriate statistical analysis for secondary outcomes. Empirical evidence and/or theory suggests that these outcomes are necessary for, or are associated with, attaining the primary outcomes. These issues may be especially important to consider for program adoption by practitioners or researchers engaged in replication work. For example, an intervention might be designed to target challenging behavior in the classroom by including a component designed to increase the number of times praise is delivered and improve the specificity/quality of the praise. According to this example, reducing challenging behavior in the classroom is the primary outcome, and increased quality and quantity of teacher praise is the secondary outcome necessary or associated with achieving the primary outcome. Distinguishing between Primary and Secondary Outcomes can become difficult. To simplify the distinction, coders should rely on the explanation provided by the authors regarding the mechanism through which change in the primary outcomes will occur. If the authors do not provide such an explanation, secondary outcomes will not be coded. Secondary Outcomes will be coded along the same dimensions as Primary Outcomes.  

C4. Percentage of secondary outcomes that are statistically significant  

C5. Overall Summary of Questions Investigated  

C5.1 Main effect analyses. Reviewers will note whether main effect questions were investigated in the study (e.g., the average effect of participation in the intervention/prevention program on primary and
C5.2 Moderator effect analyses. Reviewers will note whether the authors investigated subgroup effects (e.g., did the intervention/prevention program produce stronger effects for certain subgroups). Reviewers will specify the results of those analyses on the coding sheet when applicable.

C5.3 Mediator analyses. Reviewers will note whether the authors investigated factors or pathways that mediate program outcomes. Mediator analyses help explain the link between program activities and program outcomes (see Reynolds, 2000 for additional information on mediator analyses). Reviewers will specify the results of those analyses on the coding sheet when applicable.

Table 3

ES Indexes and Their Values for Small, Medium, and Large Effects

<table>
<thead>
<tr>
<th>Test</th>
<th>ES index</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ( m_A ) vs. ( m_B ) for independent means</td>
<td>( d = \frac{m_A - m_B}{\sigma} )</td>
<td>0.20 0.50 0.80</td>
</tr>
<tr>
<td>2. Significance of product moment r</td>
<td>( R )</td>
<td>0.10 0.30 0.50</td>
</tr>
<tr>
<td>3. ( r_A ) vs. ( r_B ) for independent rs</td>
<td>( q = z_A - z_B ) where ( z = \text{Fisher's} \ z )</td>
<td>0.05 0.15 0.25</td>
</tr>
<tr>
<td>4. ( P = 0.5 ) and the sign test</td>
<td>( g = P - 0.50 )</td>
<td>0.20 0.50 0.80</td>
</tr>
<tr>
<td>5. ( P_A ) vs. ( P_B ) for independent proportions</td>
<td>( h = \phi_A - \phi_B ) where ( \phi = \text{arc sine transformation} )</td>
<td>0.10 0.30 0.50</td>
</tr>
<tr>
<td>6. Chi-square for goodness of fit and contingency</td>
<td>( W = \sqrt{\sum_{i=1}^{k} \left( \frac{P_{i} - P_{0i}}{P_{0i}} \right)^2} )</td>
<td>0.10 0.25 0.40</td>
</tr>
<tr>
<td>7. One-way analysis of variance</td>
<td>( f = \frac{\sigma_m}{\sigma} )</td>
<td>0.02 0.15 0.35</td>
</tr>
<tr>
<td>8. Multiple and multiple partial correlation</td>
<td>( f^2 = \frac{R^2}{1 - R^2} )</td>
<td></td>
</tr>
</tbody>
</table>

Table 4

**Formulas to Compute Various Measures of Effect Size**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Effect Size</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>d index</td>
<td>( \frac{2t}{\sqrt{df}} )</td>
</tr>
<tr>
<td>F</td>
<td>f index</td>
<td>( \sqrt{\frac{df_{effect}(F)}{df_{error}}} )</td>
</tr>
<tr>
<td>_2</td>
<td>_ index</td>
<td>( \sqrt{\frac{\chi^2}{N}} )</td>
</tr>
<tr>
<td>r</td>
<td>r index</td>
<td>( r = \frac{t}{\sqrt{t^2 + df}} )</td>
</tr>
</tbody>
</table>


Table 5

**Ranges for Effect Size Indexes**

<table>
<thead>
<tr>
<th>Effect Size</th>
<th>Small</th>
<th>Medium</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>d index</td>
<td>.20</td>
<td>.50</td>
<td>.80</td>
</tr>
<tr>
<td>f index</td>
<td>.10</td>
<td>.25</td>
<td>.40</td>
</tr>
<tr>
<td>r index</td>
<td>.10</td>
<td>.30</td>
<td>.50</td>
</tr>
<tr>
<td>_ index</td>
<td>.10</td>
<td>.30</td>
<td>.50</td>
</tr>
</tbody>
</table>


Table 6

**Formulas to Convert Various Effect Size Estimates into d Index Values**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>( d = \frac{2t}{\sqrt{df_{error}}} )</td>
</tr>
<tr>
<td>F^a</td>
<td>( d = 2\sqrt{\frac{F}{df_{error}}} )</td>
</tr>
<tr>
<td>r</td>
<td>( d = \frac{2r}{\sqrt{1 - r^2}} )</td>
</tr>
<tr>
<td>_2b</td>
<td>( d = \frac{2\sqrt{\frac{\chi^2}{N} / \sqrt{1 - \chi^2 / N}}} )</td>
</tr>
</tbody>
</table>

D. Educational/Clinical Significance

Clinical significance is defined as changes in an individual’s behavior that are meaningful and important (Kazdin, 1977). Furthermore, changes in behavior must “resolve or significantly ameliorate the problem for which the client sought treatment” (Kazdin, 1977, p. 427). Clinical significance includes four domains that should be considered when evaluating the quality of evidence related to educational/clinical significance.

D1. Categorical diagnosis data. Criteria may be provided regarding inclusion into the study, or changes in diagnostic criteria at pre, post, and follow-up.

D2. Outcomes assessed via continuous variables. This criterion is defined as the percentage of participants who fall inside the range of dysfunction, the percentage of participants who fall in the normal range, or the percentage of participants showing clinical improvement during each phase of the intervention.

D3. Subjective evaluation. Reviewers should code any information provided regarding the importance of the behavioral change at pre, post, and follow up as evaluated by individuals in direct contact with the participant.

D4. Social comparison. Reviewers should code any information relating to the participant’s behavior at pre, post, and follow up in comparison to normative data such as typical peers.

To receive a rating of 3, strong evidence, a study must provide evidence in support of the clinical significance for at least 3 of the 4 criteria listed (i.e., Categorical Diagnosis Data, Outcomes Assessed Via Continuous Variables, Subjective Evaluation, or Social Comparison) during either post or follow up phases for the majority of participants.

To receive a rating of 2, promising evidence, a study must provide evidence in support of the clinical significance for at least 2 of the 4 criteria listed during either post or follow up phases for the majority of participants.

To receive a rating of 1, weak/marginal evidence, a study must provide evidence in support of the clinical significance for at least 1 of the 4 criteria for the majority of participants.

A rating of 0 indicates that none of the above criteria were met.

E. Identifiable Components

Identifiable components focus on the degree to which the authors of a study have identified unique aspects of the program/intervention that can be linked to statistically significant primary outcomes. Some interventions include multiple components. In these cases, it must be demonstrated that each component is necessary for the primary outcome to occur. For example, if an intervention designed to prevent teen smoking included a knowledge component, a peer component, and a parent component, then the study should demonstrate that all three of these components were necessary for the program to effectively prevent teen smoking. An appropriate method of demonstrating this link would be to design a study that examined the effects of the intervention with all three components, compared to the effects of the same intervention approach with only one or two of the components (e.g., peer and parent components only, knowledge and peer components only, and knowledge and parent components only). Another example of how to examine the effects of specific components might be intervention research where different approaches are compared (e.g., discussion only vs. behavioral training and discussion). As part of this code, reviewers should note the number of program components when that information is available.
Clearly this is a massive undertaking, and the Task Force recognizes that these types of investigations are rare and that the majority of studies will not have addressed this issue. Nevertheless, this process of establishing evidence linking specific intervention components to primary outcomes is critical information that can be used by practitioners and schools in making decisions about whether to implement a particular program, or portions of a particular program. Therefore, this criterion has been included, not because there are a number of studies that necessarily satisfy it currently, but because the Task Force believes this is an important direction for future intervention research.

To receive a rating of 3, strong evidence, a study must: (1) demonstrate strong evidence for significant primary outcomes, (2) use a design that allows for an analysis that identifies specific components, and (3) the analysis must provide evidence that all identified intervention components were necessary to produce change in the primary outcomes.

To receive a rating of 2, promising evidence, a study must: (1) demonstrate promising evidence for significant primary outcomes, (2) use a design that allows for an analysis which identifies specific components, and (3) the analysis must provide evidence that at least 50% of the identified intervention components were necessary to produce change in the primary outcomes.

To receive a rating of 1, weak evidence, a study must: (1) demonstrate weak evidence for significant primary outcomes, (2) use a design that allows for an analysis which identifies specific components, and (3) the analysis must provide evidence that at least 25% of the identified intervention components were necessary to produce change in the primary outcomes.

A rating of 0 indicates that there was no evidence of which components were necessary to produce change.

Additional CRITERIA FOR JUDGING IDENTIFIABLE COMPONENTS (coded descriptively):

E5. Documenting essential components. The procedures for insuring delivery of essential components are described in detail, including the procedures for training facilitators, delivery the intervention, and monitoring the intervention.

E6. Documenting adaptation. The procedures for adapting the intervention to fit the ecology (context, culture, participants) are described in detail, including documentation of the impetus (evidence-based occurrence or feedback) for adaptations and the resulting procedural changes. Each adaptation must be clearly linked to data generated by documentation techniques.

E7. Documenting contextual features. A detailed description of the context within which the intervention occurred is provided. Context refers to the community in which the intervention takes place (e.g., city, neighborhood); the specific site of implementation (e.g., school); and the specific location (e.g., classroom). Contextual variables include geographic location, population (size, demographics), social-cultural variables (e.g., relevant cultural practices or norms, social issues), and resources relevant to the intervention (e.g., necessary equipment and supplies, expenditures).

F. Implementation Fidelity

Reviewers will consider program implementation fidelity/integrity. Although there may be slight variations across implementations, data on the program integrity will be critical to determine if the intervention was responsible for the positive outcomes reported by the researchers. Two issues are considered critical here: acceptable adherence and intervention manuals. Reviewers will code whether the authors have confirmed
program implementation integrity.

To receive a rating of 3, **strong evidence**, the study must demonstrate strong evidence of acceptable adherence. In addition, evidence should be measured through at least two of the following: ongoing supervision/consultation, coding sessions, or audio/video tapes, and use of a manual. To be considered a “manual” for a rating of 3, information must have been provided to the implementers using either: (1) written materials involving a detailed account of the exact procedures and the sequence in which they are to be used or (2) a formal training session that includes a detailed account of the exact procedures and the sequence in which they are to be used. If the intervention is to be administered in “sessions” or “lessons,” then this information must be provided on a session to sessions/lesson to lesson basis. Finally, if adaptation occurs to fit varying contexts, there is a description of the procedures for adaptation.

To receive a rating of 2, **promising evidence**, the study must demonstrate evidence of acceptable adherence. In addition, evidence should be measured through at least one of the above criteria and use of a manual. To be considered a “manual” for a rating of 2, information must have been provided to the implementers using either: (1) written materials involving an overview of broad principles and a description of the intervention phases, or (2) a formal or informal training session involving an overview of broad principles and a description of the intervention phases.

To receive a rating of 1, **weak evidence**, the study must demonstrate evidence of acceptable adherence measured through at least one of the above criteria or use of a manual.

A rating of 0, no evidence, would reflect that nothing was done to ensure implementation fidelity or evidence indicates unacceptable adherence.

**F1. Acceptable adherence.**  The reviewers must be persuaded that the intervention in question was delivered in the study in a manner consistent with the intervention developer’s manual or protocol. Adequate adherence to the intervention developer’s protocol must be shown in all reviewed studies: supporting studies, null studies (i.e., those showing no difference between the candidate intervention and another intervention or control), and refuting studies (i.e., those showing the candidate intervention to be worse than another intervention or control). Acceptable adherence can be met in any one of the following ways:

**F1.1** When training, ongoing supervision and case consultation are provided by the intervention development team or another expert in that intervention.

**F1.2** When coding of sessions (live, video, or audio) has been done, it must demonstrate that, on average, the majority of authors’ or intervention developers’ criteria for adherence were satisfied.

**F1.3** When an intervention is delivered via audio- or videotapes, with or without discussion afterwards.

**F2. Manualization.** The candidate intervention should be “manualized” (i.e., accompanied by a clear description of the procedures used) and the studies must be conducted with intervention manuals and/or detailed procedural specification. Manuals may vary in form, depending on the intervention; some manuals may involve a detailed session-by-session account of exact procedures and the sequence in which they are to be used. Others may emphasize broad principles, describe phases of intervention, and offer examples of how the broad principles may be implemented. In some cases (e.g., spelling intervention program), the principles and procedures are straightforward enough that they can be described in the Method section of an article, with no need for a separate manual. In each instance, the test of whether manualization is adequate is whether the intervention procedures are sufficiently well documented that readers know what
intervention procedures were tested, and (b) intervention agents / therapists / consultants / instructors and
researchers who wish to replicate the intervention could do so faithfully (recognizing that supplemental
training may be required for nuances of the intervention to be understood). This documentation must either
be present in the article or the documentation (manual and related materials) and must be available to the
interested reader / clinician / researcher.

F3. Adaptation. The study describes the principles and procedures for adapting the intervention to
fit varying contexts, cultures and participants while maintaining essential components. Principles of
adaptation are evidence-based and derived from repetition of the intervention across multiple contexts,
cultures, and samples of participants. The procedural guide advocates repeated evaluation for each
application of the intervention.

G. Replication

Another criterion reviewers will take into account is replication. The number of replications of the
program will be coded. Replication will take into account three issues: the same intervention, the same
target problem, and replication independent of program author group.

G1. Same intervention. Two versions of an intervention program are considered to be the same if
(1) the authors judge the intervention to be essentially the same (declared in the published study, via
personal contact, or use of the same manual/procedures), and (2) the intervention duration is equal, or the
shorter version is at least 75% of the longer version, or a longer intervention is used. If neither, or only one,
of these criteria is satisfied, then the two intervention programs are considered different interventions.
Whenever there is a question about whether the intervention is the same (e.g., manual was revised for
group vs. individual administration), the reviewers will attempt to retrieve all manuals and related materials
from the author(s), review them, and come to a consensus on the same intervention issue.

G2. Same target problem. There must be at least two studies (including independent replications
within a single study) meeting all criteria treating the same target issue/problem, and the same target
age/grade group. First, studies must provide sufficient description of the target issue/problem of the treated
group to permit clear identification of the target issue/problem and target age/grade range to which the
intervention is relevant. Second, a study is considered a replication only when the intervention is applied to
a sample of the same target issue/problem and age/grade range as in the prior study(ies); otherwise, the
new study represents a new test with a new target issue/problem and/or age/grade group. When an
intervention is tested with a different target issue/problem or with a different age/grade group than in
previous research, this is considered evidence on a new application of the intervention, not a replication.

The studies will be grouped by age in the following way: infant (birth to 2 years), preschool (2-5
years), elementary (6-11 years), secondary (12-17 years), adult (18-65 years), and geriatric (over 65 years).
At transitional points (i.e., 6-11 years), reviewers will present information and recommendation for whether
to consider them the same population to the sub-domain co-chairs for a vote. Some examples of target
issue/problem to clarify: if an intervention has two studies supporting its efficacy, one intervening with
reading and the other math, this intervention does not qualify as a replication; if an intervention is supported
by two separate studies, one treating conduct problems and oppositional behavior in children 8-11 years,
and the other treating conduct disorder in children 8-11 years, this intervention may be considered
replicated, assuming it meets all other criteria. When there is question about whether a study represents a
replication with the same or different target issue/problem, the evidence will be presented to a subcommittee
and will be voted upon by the committee (see Section 3 Procedures for Decision-Making by Task Force).

G3. Independent evaluation. Reviewers will take into account any relationship between the
evaluator/researcher and the intervention program. Generally, evaluations of an intervention conducted by
the program developers tend to show larger effects than those conducted by independent evaluators. Information pertaining to an independent evaluation will be coded and noted in all reviews.

To receive a rating of **3, strong evidence**, the study must be a replication of the same intervention and target problem, be implemented by an independent evaluator, and demonstrate similar or better outcomes.

To receive a rating of **2, promising evidence**, the study must contain two of the three coding criteria (i.e., same intervention, same target problem, independent evaluation), and demonstrate similar or better outcomes.

To receive a rating of **1, weak evidence**, the study must contain at least one of these coding criteria, and demonstrate similar or better outcomes.

A rating of **0, no evidence**, none of these elements were present, and/or positive outcomes were not demonstrated.

**H. Site of Implementation**

Reviewers will note where the intervention took place. In the rating, preference is given to **school settings**, although it is recognized that interventions take place in an appropriate field-based site (e.g., home, university clinic, summer program, outpatient hospital, partial inpatient or day intervention program, inpatient hospital, private practice, mental health center, or residential intervention facility). One purpose of this information is to indicate whether the intervention took place in (a) a school setting, (b) a clinical setting where mental health services are routinely delivered independently of the study (e.g., a community mental health center, HMO, inpatient psychiatric unit), or (c) a setting specially arranged for research (e.g., university lab clinic, school classroom).

To receive a rating of **3, strong evidence**, the study must have been conducted in a public school or an alternative school, as these are where the majority of children needing intervention are located. If the type of school is unknown, a rating of 2 is the highest possible rating. In addition, interventions focusing on home-school partnerships would also receive this rating if the school initiated the intervention in an outreach effort and if the school was a public or alternative school (e.g., a public school initiated home-based program).

To receive a rating of **2, promising evidence**, the study must have been conducted in a private, charter, or university-affiliated school setting. These settings all contain special variables, which may not exist in public school settings and that may contribute to intervention effectiveness. In addition, interventions that focus on home-school partnerships would also receive this rating if the school initiated the intervention in an outreach effort and if the school was a private, charter, or university-affiliated school setting.

To receive a rating of **1, weak evidence**, the intervention was not implemented in a school setting, but could be implemented with little modification (i.e., a video parent training program that requires videotape vignettes be shown to a group of parents, followed by a psychologist facilitated discussion could be done within a school context with little modification).

A rating of **0, no evidence**, would be an intervention not implemented within a school context and would require major modifications to do so.
I. Follow Up Assessment

Reviewers will code the extent to which follow up assessment was built into the study design. The rating is based on information the authors provide on the timing of outcome assessments, the number of participants included in the follow up assessment, in addition to the consistency of assessment method used.

To receive a rating of 3, strong evidence, the study must have conducted follow up assessments over multiple intervals (e.g., 6 months, 1 year), with all participants that were included in the original sample, using similar measures used to analyze data from primary or secondary outcomes.

To receive a rating of 2, promising evidence, the study must have conducted follow up assessments at least once (e.g., 6 months), with the majority of participants that were included in the original sample, using similar measures used to analyze data from primary or secondary outcomes.

To receive a rating of 1, weak evidence at least once (e.g., 6 months), with some participants from the original sample.

A rating of 0, no evidence, would indicate that no follow up assessment was built into the study.

III. Other Descriptive or Supplemental Criteria to Consider

A. External Validity Indicators

A1. Participant selection (sampling). Participants refer to all persons involved in the intervention, including recipients (e.g., students), implementers (e.g., teachers, psychologists), staff trainers (e.g., researchers, interventionists), decision makers (e.g., principal, research team members), and other key stakeholders who may influence the other participants (e.g., parents, community members, superintendent). Participants are described in sufficient detail to permit other researchers and interventionists to determine both the conditions under which the intervention occurred and the generalizability to intended participants. Reviewers will consider any information reported by the authors regarding the sources from which the sample was drawn, and the means by which clients came to be in the study (e.g., screening at an elementary school, university clinic pool of families, newspaper advertisements, community mental health referral).

A1.1 Inclusion/exclusion criteria specified. Reviewers will consider what participant selection criteria the authors applied to the sample that might limit its representativeness (e.g., whether authors required parent participation, required a two-parent family, or whether they excluded comorbid cases).

A1.2. Inclusion/exclusion criteria similar to school practice. Reviewers will also consider whether inclusion/exclusion criteria were similar to those likely to be used in school practice (code as yes/no). For example, criteria that are similar to school practice might include ruling several physically aggressive students out of a prevention program, and excluding youth from individual psychotherapy who are currently abusing substances. Criteria that are dissimilar from school practice might involve, for example, excluding children who do not have two parents living at home, or eliminating cases with comorbid diagnoses other than the one targeted in intervention.
A1.3 Specified criteria related to concern. The criteria for inclusion into the study should be related to the goal of the intervention. For example, if an intervention is designed to reduce aggressive behavior, then some specified criterion related to the presence of aggressive behavior should be included in the study.

A2. Participant characteristics specified for treatment and control group. Characteristics of the samples used to test interventions will be specified. An objective here is to identify the range of client (e.g., child/family/mediator) characteristics within which an intervention has been shown to have beneficial effects. Each intervention and control sample will be described along the following demographic dimensions:

A2.1 Type of participant. Specify student, parent, teacher, or larger system
A2.2 Grade/age
A2.3 Gender
A2.4 Ethnicity or multiethnic
A2.5 Ethnic identity
A2.6 Race(s)
A2.7 Acculturation
A2.8 Primary language
A2.9 SES
A2.10 Family structure (e.g., single parent, immigrant status)
A2.11 Locale (e.g., urban, rural, suburban, university-affiliated site)
A2.12 Disability (or other special considerations such as English Language Learners)
A2.13 Functional descriptors (e.g., clinical, sub-clinical, "normal" groups; other academic problems present, such as at-risk factors, low reading achievement, etc.)

A3. Details are provided regarding demographic variables (age, gender, ethnicity-race) and any other variables that have been identified as:

A3.1. Having differential relevance for intended outcomes.
A3.2. Being relevant to inclusion or exclusion (e.g., level of education, prior experience).

A4. Receptivity/acceptance by target participant population. Please note any consumer satisfaction information, such as reactions from children, parents, teachers, or other program participants as well as reactions to and perceptions of what was gained as a result of the program. Reviewers should also provide a global rating as to whether the participants reported benefiting or not benefiting from the intervention.

A5. Generalization of effects. The extent to which the outcomes generalize across time, settings, and persons. Generalization must be documented through systematic data collection.

A5.1 Generalization over time.

A5.1.1 Persistence (sustainability) of outcomes. Evidence is provided regarding the extent to which outcomes persist following the termination of the intervention.

A5.1.2 Procedures for ensuring sustainability. The conditions under which
sustainability has been achieved; that is, documentation of efforts (or lack thereof) to ensure maintenance of outcomes (e.g., through booster sessions).

A5.2 Generalization across settings.

A5.2.1 Application outside of the intervention context. Evidence is provided regarding the extent to which outcomes are manifested in contexts that are different from the intervention context. For example, evidence of the capacity for the participants to make use of newly acquired skills not only within the intervention context (e.g., resolving conflicts with peers in the classroom) but also in other settings (e.g., on the playground) or applied to other types of problems (e.g., solving conflicts with adults).

A5.2.2 Procedures for ensuring application. The conditions under which application outside of the intervention context was achieved; that is, documentation of efforts (or lack thereof) to ensure application to other settings, problems, etc. (e.g., through simulated practice in wide scale application; booster sessions in other contexts).

A5.2.3 Institutionalization. The extent to which impact on implementers or context is sustained, documented through systematic follow-up. For example, the extent to which teachers continue the application of a classroom-based intervention after the end of the formal intervention program; or the extent to which the school system continues discipline practices established by the intervention program. The conditions under which sustainability has been achieved; that is, documentation of efforts (or lack thereof) to ensure sustained impact (e.g., through booster sessions or follow-up consultation).

A5.3 Generalization across persons

Evidence is provided regarding the extent to which outcomes are manifested with participants who are different than the original group of participants for which the intervention was evaluated. An underlying assumption is that each implementation of an intervention in a different context or with a different group constitutes a “different” intervention, given the cultural-contextual nature of interventions.

B. Length of Intervention

Reviewers will code the length of the intervention (i.e., how long the intervention was in place) in weeks, months, or years.

C. Intensity of Intervention

Intensity refers to the length of the intervention sessions and the frequency of the sessions. Reviewers will code both length and frequency of intervention sessions.

D. Dosage/Response

Reviewers will note whether there is evidence that length or intensity of intervention was associated with stronger outcomes or better performance on outcome measures.

E. Program Implementer

Reviewers will code who conducted the intervention. Possible choices include: research staff, school
specialty staff (e.g., counselors, psychologists, social workers, special educators, etc.), teachers, educational assistants, parents, college students, peers, and others. If research staff were only involved in training intervention implementers, then do not include them as program implementers, unless the intervention is the training. This support will be documented under Training and Support Resources.

F. Characteristics of the Intervener

Reviewers will note how similar the target participants were to the intervener on several characteristics (e.g., race, gender, SES).

G. Intervention Style or Orientation

Reviewers will note the theoretical underpinnings of the intervention program, such as behavioral, cognitive-behavioral, experiential, humanistic/interpersonal, psychodynamic/insight oriented, or combination/other.

H. Cost Analysis Data

Cost analysis data provided (coded yes/no). Whenever authors report data regarding the cost to implement the program, and/or cost-to-benefit analyses, reviewers will report this information.

I. Training and Support Resources

What supports were provided if school or other typical staff implemented the intervention in the study?

I1. Simple orientation given to change agents

I2. Training workshops given (indicate # of workshops, average length of training, and who conducted the training)

I3. On-going technical support provided once intervention began, such as phone or face-to-face consultation, classroom observations, coaching or trouble-shooting after program has begun

I4. Program materials must be obtained to conduct intervention

I5. Special facilities (extra space) must be obtained

I6. Other

J. Feasibility

Reviewers will rate the level of difficulty in training intervention agents (i.e., high, moderate, or low) and report and rate the cost involved in training intervention agents (i.e., high, moderate, low) when this information is available.
Domain:

☐ School- and community-based intervention programs for social and behavioral problems
☐ Academic intervention programs
☐ Family and parent intervention programs
☐ School-wide and classroom-based programs
☐ Comprehensive and coordinated school health services

Name of Coder(s): ____________________________ Date: ____________________________

M / D / Y

Full Study Reference in APA format: __________________________________________________________
________________________________________________________________________________________
________________________________________________________________________________________

Intervention Name (description from study): ____________________________________________________
________________________________________________________________________________________
________________________________________________________________________________________

Study ID Number (Unique Identifier): __________________________________________________________

Type of Publication:  (Check one)

☐ Book/Monograph
☐ Journal article
☐ Book chapter
☐ Other (specify):
I. General Characteristics

A. General Design Characteristics

A1. Random assignment designs (if random assignment design, select one of the following)
   A1.1 □ Completely randomized design
   A1.2 □ Randomized block design (between-subjects variation)
   A1.3 □ Randomized block design (within-subjects variation)
   A1.4 □ Randomized hierarchical design

A2. Nonrandomized designs (if nonrandom assignment design, select one of the following)
   A2.1 □ Nonrandomized design
   A2.2 □ Nonrandomized block design (between-participants variation)
   A2.3 □ Nonrandomized block design (within-participants variation)
   A2.4 □ Nonrandomized hierarchical design
   A2.5 □ Optional coding of Quasi-experimental designs (see Appendix C)

A3. Overall confidence of judgment on how participants were assigned (select one of the following)
   A3.1 □ Very low (little basis)
   A3.2 □ Low (guess)
   A3.3 □ Moderate (weak inference)
   A3.4 □ High (strong inference)
   A3.5 □ Very high (explicitly stated)
   A3.6 □ N/A
   A3.7 □ Unknown/unable to code

B. Statistical Treatment/Data Analysis (answer B1 through B6)

   B1. Appropriate unit of analysis □yes □no
   B2. Familywise error rate controlled □yes □no □ N/A
   B3. Sufficiently large N □yes □no
      Statistical Test: __________
      level: __________
      ES: __________
      N required: __________

   B4. Total size of sample (start of the study): __________
      N
   B5. Intervention group sample size: __________
      N
   B6. Control group sample size: __________
      N

For studies using qualitative research methods, code B7 and B8

B7. Coding

   B7.1 Coding scheme linked to study’s theoretical-empirical basis (select one) □yes □no

   B7.2 Procedures for ensuring consistency of coding are used (select one) □yes □no
      Describe procedures: ________________________________
B7.3 Progression from abstract concepts to empirical exemplars is clearly articulated (select one)  □ yes  □ no

B8. Interactive process followed (select one)  □ yes  □ no

Describe process: _______________________________

C. Type of Program (select one)

C1. □ Universal prevention program
C2. □ Selective prevention program
C3. □ Targeted prevention program
C4. □ Intervention/Treatment
C5. □ Unknown

D. Stage of the Program (select one)

D1. □ Model/demonstration programs
D2. □ Early stage programs
D3. □ Established/institutionalized programs
D4. □ Unknown

E. Concurrent or Historical Intervention Exposure (select one)

E1. □ Current exposure
E2. □ Prior exposure
E3. □ Unknown
II. Key Features for Coding Studies and Rating Level of Evidence/ Support

(3=Strong Evidence  2=Promising Evidence  1=Weak Evidence  0=No Evidence)

A. Measurement (answer A1 through A4)

   A1. Use of outcome measures that produce reliable scores for the majority of primary outcomes. The table for Primary/Secondary Outcomes Statistically Significant allows for listing separate outcomes and will facilitate decision making regarding measurement (select one of the following)

   - A1.1 Yes
   - A1.2 No
   - A1.3 Unknown/unable to code

   A2. Multi-method (select one of the following)

   - A2.1 Yes
   - A2.2 No
   - A2.3 N/A
   - A2.4 Unknown/unable to code

   A3. Multi-source (select one of the following)

   - A3.1 Yes
   - A3.2 No
   - A3.3 N/A
   - A3.4 Unknown/unable to code

   A4. Validity of measures reported (select one of the following)

   - A5.1 Yes validated with specific target group
   - A5.2 In part, validated for general population only
   - A5.3 No
   - A5.4 Unknown/unable to code

   Rating for Measurement (select 0, 1, 2, or 3):  3  2  1  0

B. Comparison Group

   B1. Type of Comparison Group (select one of the following)

   - B1.1 Typical contact
   - B1.2 Typical contact (other) specify:
   - B1.3 Attention placebo
   - B1.4 Intervention elements placebo
   - B1.5 Alternative intervention
   - B1.6 PharmacotherapyB1.1
   - B1.7 No intervention
   - B1.8 Wait list/delayed intervention
   - B1.9 Minimal contact
   - B1.10 Unable to identify comparison group

   Rating for Comparison Group (select 0, 1, 2, or 3):  3  2  1  0
B2. Overall confidence rating in judgment of type of comparison group (select one of the following)

- B2.1 □ Very low (little basis)
- B2.2 □ Low (guess)
- B2.3 □ Moderate (weak inference)
- B2.4 □ High (strong inference)
- B2.5 □ Very high (explicitly stated)
- B2.6 □ Unknown/Unable to code

B3. Counterbalancing of Change Agents (answer B3.1 to B3.3)

- B3.1 □ By change agent
- B3.2 □ Statistical
- B3.3 □ Other

B4. Group Equivalence Established (select one of the following)

- B4.1 □ Random assignment
- B4.2 □ Posthoc matched set
- B4.3 □ Statistical matching
- B4.4 □ Post hoc test for group equivalence

B5. Equivalent Mortality (answer B5.1 through B5.3)

- B5.1 □ Low Attrition (less than 20% for Post)
- B5.2 □ Low Attrition (less than 30% for follow-up)
- B5.3 □ Intent to intervene analysis carried out

Findings

C. Primary/Secondary Outcomes Are Statistically Significant

C1. Evidence of appropriate statistical analysis for primary outcomes (answer C1.1 through C1.3)

- C1.1 □ Appropriate unit of analysis (rate from previous code)
- C1.2 □ Familywise/experimenterwise error rate controlled when applicable (rate from previous code)
- C1.3 □ Sufficiently large N (rate from previous code)

C2. Percentage of primary outcomes that are significant (select one of the following)

- C2.1 □ Significant primary outcomes for at least 75% of the total primary outcome measures for each key construct
- C2.2 □ Significant primary outcomes for between 50% and 74% of the total primary outcome measures for each key construct
- C2.3 □ Significant primary outcomes for between 25% and 49% of the total primary outcome measures for any key construct

Rating for Primary Outcomes Statistically Significant (select 0, 1, 2, or 3): □ 3 □ 2 □ 1 □ 0

C3. Evidence of appropriate statistical analysis for secondary outcomes (answer C3.1 through C3.3)

- C3.1 □ Appropriate unit of analysis
- C3.2 □ Familywise/experimenterwise error rate controlled when applicable (rate from previous code)
C3.3 □ Sufficiently large N (rate from previous code)

C4. Percentage of secondary outcomes that are significant (select one of the following)

C4.1 □ Significant secondary outcomes for at least 75% of the total secondary outcome measures for each key construct

C4.2 □ Significant secondary outcomes for between 50% and 74% of the total secondary outcome measures for each key construct

C4.3 □ Significant secondary outcomes for between 25% and 49% of the total secondary outcome measures for any key construct

Rating for Secondary Outcomes Statistically Significant (select 0, 1, 2, or 3): [ ] 3 [ ] 2 [ ] 1 [ ] 0

C5. Overall Summary of Questions Investigated

C5.1 Main effect analyses conducted (select one) □ yes □ no
Specify results: _________________________________________________________

C5.2 Moderator effect analyses conducted (select one) □ yes □ no
Specify results: _________________________________________________________

C5.3 Mediator analyses conducted (select one) □ yes □ no
Specify results: _________________________________________________________
### C. Primary/Secondary Outcomes Statistically Significant (only list \( p \leq .05 \))

(list primary outcomes first in alphabetical order, followed by secondary outcomes in alphabetical order)

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Primary vs. Secondary</th>
<th>Who Changed</th>
<th>What Changed</th>
<th>Source</th>
<th>Treatment Information</th>
<th>Outcome Measure Used</th>
<th>Reliability</th>
<th>ES (1-_)</th>
</tr>
</thead>
</table>
| Outcome #1: | □ Primary  
□ Secondary  
□ Unknown | □ Child  
□ Teacher  
□ Parent/sign. adult  
□ Ecology  
□ Other  
□ Unknown | □ Behavior  
□ Attitude  
□ Knowledge  
□ Other  
□ Unknown | □ Self Report  
□ Parent Report  
□ Teacher Report  
□ Observation  
□ Test  
□ Other  
□ Unknown |
| Outcome #2: | □ Primary  
□ Secondary  
□ Unknown | □ Child  
□ Teacher  
□ Parent/sign. Adult  
□ Ecology  
□ Other  
□ Unknown | □ Behavior  
□ Attitude  
□ Knowledge  
□ Other  
□ Unknown | □ Self Report  
□ Parent Report  
□ Teacher Report  
□ Observation  
□ Test  
□ Other  
□ Unknown |
| Outcome #3: | □ Primary  
□ Secondary  
□ Unknown | □ Child  
□ Teacher  
□ Parent/sign. Adult  
□ Ecology  
□ Other  
□ Unknown | □ Behavior  
□ Attitude  
□ Knowledge  
□ Other  
□ Unknown | □ Self Report  
□ Parent Report  
□ Teacher Report  
□ Observation  
□ Test  
□ Other  
□ Unknown |
| Outcome #4: | □ Primary  
□ Secondary  
□ Unknown | □ Child  
□ Teacher  
□ Parent/sign. Adult  
□ Ecology  
□ Other  
□ Unknown | □ Behavior  
□ Attitude  
□ Knowledge  
□ Other  
□ Unknown | □ Self Report  
□ Parent Report  
□ Teacher Report  
□ Observation  
□ Test  
□ Other  
□ Unknown |
| Outcome #5: | □ Primary  
□ Secondary  
□ Unknown | □ Child  
□ Teacher  
□ Parent/sign. Adult  
□ Ecology  
□ Other  
□ Unknown | □ Behavior  
□ Attitude  
□ Knowledge  
□ Other  
□ Unknown | □ Self Report  
□ Parent Report  
□ Teacher Report  
□ Observation  
□ Test  
□ Other  
□ Unknown |
Null Findings/Negative Outcomes Associated with the Intervention  (listed alphabetically by outcome)

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Primary vs. Secondary</th>
<th>Who Was Targeted for Change</th>
<th>What Was Targeted for Change</th>
<th>Source</th>
<th>Note null/negative outcomes</th>
<th>Outcome Measure Used</th>
<th>Reliability</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome #1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome #2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome #3:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome #4:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome #5:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Type of Data Effect Size is Based On

(check all that apply)

- Means and SDs
- t-value or F-value
- Chi-square (df = 1)
- Frequencies or proportions (dichotomous)
- Frequencies or proportions (polytomous)
- Other (specify):
- Unknown

### Confidence Rating in ES Computation

(select one of the following)

- Highly estimated (e.g., only have N p value)
- Moderate estimation (e.g., have complex but complete statistics)
- Some estimation (e.g., unconventional statistics that require conversion)
- Slight estimation (e.g., use significance testing statistics rather than descriptives)
- No estimation (e.g., all descriptive data is present)

### D. Educational/Clinical Significance

<table>
<thead>
<tr>
<th>Outcome Variables:</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Follow Up</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>D1. Categorical Diagnosis Data</strong></td>
<td>Diagnostic information regarding inclusion into the study presented:</td>
<td>Positive change in diagnostic criteria from pre to posttest:</td>
<td>Positive change in diagnostic criteria from posttest to follow up:</td>
</tr>
<tr>
<td></td>
<td>□ Yes □ No □ Unknown</td>
<td>□ Yes □ No □ Unknown</td>
<td>□ Yes □ No □ Unknown</td>
</tr>
<tr>
<td><strong>D2. Outcome Assessed via continuous Variables</strong></td>
<td>Positive change in percentage of participants showing clinical improvement from pre to posttest:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>□ Yes □ No □ Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>D3. Subjective Evaluation:</strong> The importance of behavior change is evaluated by individuals in direct contact with the participant.</td>
<td>Importance of behavior change is evaluated:</td>
<td>Importance of behavior change from pre to posttest is evaluated positively by individuals in direct contact with the participant:</td>
<td>Importance of behavior change from posttest to follow up is evaluated positively by individuals in direct contact with the participant:</td>
</tr>
<tr>
<td></td>
<td>□ Yes □ No □ Unknown</td>
<td>□ Yes □ No □ Unknown</td>
<td>□ Yes □ No □ Unknown</td>
</tr>
<tr>
<td><strong>D4. Social Comparison:</strong> Behavior of participant at pre, post, and follow up is compared to normative data (e.g., a typical peer).</td>
<td>Participant’s behavior is compared to normative data</td>
<td>Participant’s behavior has improved from pre to posttest when compared to normative data:</td>
<td>Participant’s behavior has improved from posttest to follow up when compared to normative data:</td>
</tr>
<tr>
<td></td>
<td>□ Yes □ No □ Unknown</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Rating for Educational/Clinical Significance (select 0, 1, 2, or 3): □ 3 □ 2 □ 1 □ 0

### E. Identifiable Components (answer E1 through E7)

- **E1. Evidence for primary outcomes (rate from previous code):** □ 3 □ 2 □ 1 □ 0
- **E2. Design allows for analysis of identifiable components (select one):** □ yes □ no

**E3. Total number of components:** N
E4. Number of components linked to primary outcomes: \( N \)

Additional criteria to code descriptively:

E5. Clear documentation of essential components (select one) \( \square \) yes \( \square \) no

E6. Procedures for adapting the intervention are described in detail (select one) \( \square \) yes \( \square \) no

E7. Contextual features of the intervention are documented (select one) \( \square \) yes \( \square \) no

Rating for Identifiable Components (select 0, 1, 2, or 3): \( \square 3 \) \( \square 2 \) \( \square 1 \) \( \square 0 \)

F. Implementation Fidelity

F1. Evidence of Acceptable Adherence (answer F1.1 through F1.3)

\( \square \) Ongoing supervision/consultation
\( \square \) Coding intervention sessions/lessons or procedures
\( \square \) Audio/video tape implementation (select F1.3.1 or F1.3.2):

\( \square \) Entire intervention
\( \square \) Part of intervention

F2. Manualization (select all that apply)

\( \square \) Written material involving a detailed account of the exact procedures and the sequence in which they are to be used
\( \square \) Formal training session that includes a detailed account of the exact procedures and the sequence in which they are to be used
\( \square \) Written material involving an overview of broad principles and a description of the intervention phases
\( \square \) Formal or informal training session involving an overview of broad principles and a description of the intervention phases

F3. Adaptation procedures are specified (select one) \( \square \) yes \( \square \) no \( \square \) unknown

Rating for Implementation Fidelity (select 0, 1, 2, or 3): \( \square 3 \) \( \square 2 \) \( \square 1 \) \( \square 0 \)

G. Replication (answer G1, G2, G3, and G4)

\( \square \) Same Intervention
\( \square \) Same Target Problem
\( \square \) Independent evaluation

Rating for Replication (select 0, 1, 2, or 3): \( \square 3 \) \( \square 2 \) \( \square 1 \) \( \square 0 \)

H. Site of Implementation

H1. School (if school is the site, select one of the following options)

\( \square \) Public
H1.2  □ Private  
H1.3  □ Charter  
H1.4  □ University Affiliated  
H1.5  □ Alternative  
H1.6  □ Not specified/unknown  

H2. Non School Site (if it is a non school site, select one of the following options)  
H2.1  □ Home  
H2.2  □ University Clinic  
H2.3  □ Summer Program  
H2.4  □ Outpatient Hospital  
H2.5  □ Partial inpatient/day Intervention Program  
H2.6  □ Inpatient Hospital  
H2.7  □ Private Practice  
H2.8  □ Mental Health Center  
H2.9  □ Residential Treatment Facility  
H2.10  □ Other (specify): ____________________________  
H2.11  □ Unknown/insufficient information provided  

Rating for Site of Implementation (select 0, 1, 2, or 3):  □ 3  □ 2  □ 1  □ 0  

I. Follow Up Assessment  

□ Timing of follow up assessment: specify________________________  

□ Number of participants included in the follow up assessment: specify________________________  

□ Consistency of assessment method used: specify________________________  

Rating for Follow Up Assessment (select 0, 1, 2, or 3):  □ 3  □ 2  □ 1  □ 0  

III. Other Descriptive or Supplemental Criteria to Consider  
A. External Validity Indicators  
A1. Sampling procedures described in detail □ yes □ no  
  
  Specify rationale for selection: __________________________________________  
  
  Specify rationale for sample size: _________________________________________  
  
  A1.1 Inclusion/exclusion criteria specified □ yes □ no  
  
  A1.2 Inclusion/exclusion criteria similar to school practice □ yes □ no  
  
  A1.3 Specified criteria related to concern □ yes □ no  

A2. Participant Characteristics Specified for Treatment and Control Group
<table>
<thead>
<tr>
<th>Participants from Treatment Group</th>
<th>Grade/age</th>
<th>Gender</th>
<th>Ethnicity or Multi-ethnic</th>
<th>Ethnic Identity</th>
<th>Race(s)</th>
<th>Acculturation</th>
<th>Primary Language</th>
<th>SES</th>
<th>Family Structure</th>
<th>Locale</th>
<th>Disability</th>
<th>Functional Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child/Student</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent/caregiver</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participants from Control Group</th>
<th>Grade/age</th>
<th>Gender</th>
<th>Ethnicity or Multi-ethnic</th>
<th>Ethnic Identity</th>
<th>Race(s)</th>
<th>Acculturation</th>
<th>Primary Language</th>
<th>SES</th>
<th>Family Structure</th>
<th>Locale</th>
<th>Disability</th>
<th>Functional Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child/Student</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent/caregiver</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A3. Details are provided regarding variables that:

A3.1 Have differential relevance for intended outcomes [ ]yes [ ]no

Specify: ____________________________
A3.2 Have relevance to inclusion criteria □yes □no

Specify: ____________________________

A4. Receptivity/acceptance by target participant population (treatment group)

<table>
<thead>
<tr>
<th>Participants from Treatment Group</th>
<th>Results (What person reported to have gained from participation in program)</th>
<th>General Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Child/Student</td>
<td>□ Participants reported benefiting overall from the intervention</td>
<td></td>
</tr>
<tr>
<td>□ Parent/caregiver</td>
<td>□ Participants reported not benefiting overall from the intervention</td>
<td></td>
</tr>
<tr>
<td>□ Teacher</td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ School</td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Other</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A5. Generalization of Effects:

A5.1 Generalization over time

A5.1.1 Evidence is provided regarding the sustainability of outcomes after intervention is terminated □yes □no

Specify: ____________________________

A5.1.2 Procedures for maintaining outcomes are specified □yes □no

Specify: ____________________________

A5.2 Generalization across settings

A5.2.1 Evidence is provided regarding the extent to which outcomes are manifested in contexts that are different from the intervention context □yes □no

Specify: ____________________________

A5.2.2 Documentation of efforts to ensure application of intervention to other settings □yes □no
Specify: ________________________________

A5.2.3 Impact on implementers or context is sustained □yes  □no

Specify: ________________________________

A5.3 Generalization across persons

Evidence is provided regarding the degree to which outcomes are manifested with participants who are different than the original group of participants for with the intervention was evaluated □yes  □no

Specify: ________________________________

B. Length of Intervention (select B1 or B2)

B1. □ Unknown/insufficient information provided

B2. □ Information provided (if information is provided, specify one of the following:)

B2.1 weeks _____

B2.2 months _____

B2.3 years _____

B2.4 other _____

C. Intensity/dosage of Intervention (select C1 or C2)

C1. □ Unknown/insufficient information provided

C2. □ Information provided (if information is provided, specify both of the following:)

C2.1 length of intervention session _____

C2.2 frequency of intervention session _____

D. Dosage Response (select D1 or D2)

D1. □ Unknown/insufficient information provided

D2. □ Information provided (if information is provided, answer D2.1)

D2.1 Describe positive outcomes associated with higher dosage: ________________________
E. Program Implementer (select all that apply)

E1. Research Staff
E2. School Specialty Staff
E3. Teachers
E4. Educational Assistants
E5. Parents
E6. College Students
E7. Peers
E8. Other
E9. Unknown/insufficient information provided

F. Characteristics of the Intervener

F1. Highly similar to target participants on key variables (e.g., race, gender, SES)
F2. Somewhat similar to target participants on key variables
F3. Different from target participants on key variables

G. Intervention Style or Orientation (select all that apply)

G1. Behavioral
G2. Cognitive-behavioral
G3. Experiential
G4. Humanistic/interpersonal
G5. Psychodynamic/insight oriented
G6. Other (specify): ___________________
G7. Unknown/insufficient information provided

H. Cost Analysis Data (select G1 or G2)

H1. Unknown/insufficient information provided
H2. Information provided (if information is provided, answer H2.1)

H2.1 Estimated Cost of Implementation: ________________________

I. Training and Support Resources (select all that apply)

I1. Simple orientation given to change agents
I2. Training workshops conducted

# of Workshops provided _____
Average length of training _____

Who conducted training (select all that apply)

I2.1 Project Director
I2.2 Graduate/project assistants
I2.3 □ Other (please specify):
□ Unknown

I3. □ Ongoing technical support
I4. □ Program materials obtained
I5. □ Special Facilities
I6. □ Other (specify):

J. Feasibility

J1. Level of difficulty in training intervention agents (select one of the following)
  J1.1 □ High
  J1.2 □ Moderate
  J1.3 □ Low
  J1.4 □ Unknown

J2. Cost to train intervention agents (specify if known): ______________________

J3. Rating of cost to train intervention agents (select one of the following)
  J3.1 □ High
  J3.2 □ Moderate
  J3.3 □ Low
  J3.4 □ Unknown
### Summary of Evidence for Group-Based Design Studies

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Overall Evidence Rating</th>
<th>Description of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NNR = No numerical rating or 0 - 3</td>
<td>Strong Promising Weak No/limited evidence or Descriptive ratings</td>
</tr>
</tbody>
</table>

#### General Characteristics

- General Design Characteristics
- Statistical Treatment
- Type of Program
- Stage of Program
- Concurrent/Historical Intervention Exposure

#### Key Features

- Measurement
- Comparison Group
- Primary/Secondary Outcomes are Statistically Significant
- Educational/clinical significance
- Identifiable Components
- Implementation Fidelity
- Replication
- Site of Implementation
- Follow Up Assessment Conducted
<table>
<thead>
<tr>
<th>Descriptive or Supplemental Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>External validity indicators</td>
</tr>
<tr>
<td>Length of Intervention</td>
</tr>
<tr>
<td>Intensity/dosage</td>
</tr>
<tr>
<td>Dosage Response</td>
</tr>
<tr>
<td>Program Implementer</td>
</tr>
<tr>
<td>Characteristics of the Intervener</td>
</tr>
<tr>
<td>Intervention Style/Orientation</td>
</tr>
<tr>
<td>Cost Analysis Data Provided</td>
</tr>
<tr>
<td>Training and Support Resources</td>
</tr>
<tr>
<td>Feasibility</td>
</tr>
</tbody>
</table>
CODING MANUAL FOR

SINGLE-PARTICIPANT INTERVENTION RESEARCH DESIGN

Task Force on Evidence-Based Interventions in

School Psychology
The Coding Manual for Single-Participant Intervention Research Design was created by the Task Force on Evidence-Based Interventions in School Psychology.

The following individuals contributed to development of the Coding Manual:

*Sandra Christenson, PhD
University of Minnesota

*Joseph A. Durlak, PhD
Loyola University – Chicago

*Thomas R. Kratochwill, PhD
University of Wisconsin-Madison

*Joel R. Levin, PhD
University of Arizona

*Gretchen Lewis-Snyder
University of Wisconsin-Madison

*Karen Callan Stoiber, PhD
University of Wisconsin-Milwaukee

*Elisa Steele Shernoff
University of Wisconsin-Madison

*Ronda Talley, PhD, MPH
Tri-T Associates, Inc.

*Greg A. Waas, PhD
Northern Illinois University

Stacey Bates
University of Texas

Laura Guli
University of Texas

Dan Bolt, PhD
University of Wisconsin-Madison

Bonnie Nastasi, PhD
Institute for Community Research, CT

Cindy Carlson, PhD
University of Texas

Arthur Reynolds, PhD
University of Wisconsin-Madison

Diane Cox
University of Texas

Steve Schensul, PhD
University of Connecticut Health Center

Maria Fischel
University of Texas

Carmen Valdez
University of Texas

The Task Force also acknowledges contribution of the Interdisciplinary Qualitative Research Committee:

<table>
<thead>
<tr>
<th>Name</th>
<th>Specialty Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bonnie Nastasi, Co-Chair</td>
<td>School Psychology</td>
</tr>
<tr>
<td>Stephen Schensul, Co-Chair</td>
<td>Anthropology</td>
</tr>
<tr>
<td>Doug Campbell</td>
<td>Educational Research</td>
</tr>
<tr>
<td>Denise DeZolt</td>
<td>School Psychology</td>
</tr>
<tr>
<td>David Fettermann</td>
<td>Anthropology</td>
</tr>
<tr>
<td>Karen Harris</td>
<td>Special Education/Educational Psychology</td>
</tr>
<tr>
<td>Evelyn Jacob</td>
<td>Anthropology</td>
</tr>
<tr>
<td>Colette Ingraham</td>
<td>School Psychology</td>
</tr>
<tr>
<td>Margaret LeCompte</td>
<td>Sociology</td>
</tr>
<tr>
<td>Joel Meyers</td>
<td>School Psychology</td>
</tr>
<tr>
<td>Evelyn Oka</td>
<td>School Psychology</td>
</tr>
<tr>
<td>Mittie Quinn</td>
<td>School Psychology</td>
</tr>
<tr>
<td>Jean J. Schensul</td>
<td>Anthropology</td>
</tr>
<tr>
<td>Cherie Tyler</td>
<td>School Psychology</td>
</tr>
<tr>
<td>Stephen Truscott</td>
<td>School Psychology</td>
</tr>
</tbody>
</table>
* Designates membership on the original Manual Subcommittee.
• Designates graduate student

For further information please contact:

Thomas R. Kratochwill, PhD
Chair, Task Force on Evidence-Based Interventions in School Psychology
School Psychology Program
1025 W. Johnson Street
University of Wisconsin-Madison
Madison, WI 53706-1796
tomkat@education.wisc.edu
(608) 262-5912
SINGLE-PARTICIPANT DESIGN

I. General Characteristics

A. General Design Characteristics

Studies will be classified according to the type of single-participant design, given the author(s) provide sufficient information.

A1. Type of study. Specify type

A1.1 Within series. Within series designs evaluate change in client measures within various phases of the investigation. The most basic form is an A/B/A/B strategy in which A represents a baseline phase and B represents an intervention designed specifically for the participant (slashes in the design denote separate phases). In this type of design, the intervention is reintroduced during the second B phase. The tactic of repeating intervention effects can also be used in comparing different interventions. For example, a researcher can compare a B phase to a C in several sequential phases (e.g., B/C/B/C/B/C; see Table 1).

A1.2 Between-series. There are two types of between-series designs, the Alternating Treatment Design (ATD) and the Simultaneous Treatment Design (STD). With the ATD the investigator is able to compare two or more interventions in a relatively brief period of time while avoiding some of the major disadvantages of withdrawal designs (e.g., withdrawal of intervention, stability of the data series, among other features). In contrast to the ATD design, the STD presents interventions to the participants simultaneously (Kazdin & Hartmann, 1978). Nevertheless, the simultaneous availability of the interventions does not necessarily insure that the client is exposed to all interventions equally. In fact, the STD really may allow the evaluation of a client’s “preference” among interventions because the interventions are available at the same time in the same session.

It is possible that the STD could provide the researcher with information on client responsiveness to interventions where definite preferences exist. Although the STD and ATD allow the investigator to compare two or more interventions, the designs are not limited to two interventions. The researcher can schedule more than two, but the logistical considerations in balancing all the features of the designs with three or more conditions may be challenging. Thus, for practical reasons the usual applications of these designs involve only two intervention comparisons.

A1.3 Combined series/multiple baseline. In the combined series single-participant design the researcher draws a comparison both within and between series. The multiple baseline design (MBD) represents the most common example of this strategy because it includes a simple within-phase element and replicates the intervention across either participants, settings, or behaviors. The internal validity of the design is met through staggering the interventions across time. The MBD across participants is regarded as the strongest of the three design variations because the replication occurs across individual units.
Table 1. Major Types of Single-Participant Designs and Associated Characteristics

<table>
<thead>
<tr>
<th>Design Type</th>
<th>Representativeness</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within-series element</td>
<td>Simple phase change design examples: A/B, A/B/A, A/B/A/B, B/C/B/C</td>
<td>In these design elements, variability level and trend within a data series are assessed under similar conditions, the independent variable is introduced, and concomitant changes are assessed in the stability, level, and trend across phases of a single data series.</td>
</tr>
<tr>
<td></td>
<td>Complex phase change (B/B+C/B, C/B+C/C, B+C/B, B+C/B+B+C)</td>
<td></td>
</tr>
<tr>
<td>Between-series elements</td>
<td>Alternating treatment design (ATD)</td>
<td>In these design elements, estimates of variability, level, and trends in a data series are assessed on measures within a specific condition and across time. Outcomes are assessed by comparing two or more of these series.</td>
</tr>
<tr>
<td></td>
<td>Simultaneous treatment design (STD)</td>
<td></td>
</tr>
<tr>
<td>Combined-series elements</td>
<td>Multiple baseline design examples: across participants, across behaviors, across situations</td>
<td>In these design elements, comparisons are made both between and within a data series. Repetitions of a single simple phase change are scheduled, each with a new series and in which both the length and timing of the phase change differ across repetitions.</td>
</tr>
</tbody>
</table>


Once changes are observed in the first A/B series, the remaining series receive the intervention in sequential fashion. Each time a replication occurs across either participants, settings, or behaviors, the researcher is more confident that there is a causal relation between the independent and dependent variable. These designs are most useful because the researcher has the option of implementing them across three different aspects of dependent measures (i.e., across settings, behavior, and individuals). Also, the researcher can schedule further replications depending upon practical and ethical considerations.

**A1.4 Mixed designs.** A combination of single-participant and group designs or a combination of the following designs: A1.1, A1.2, A1.3 (see Coding Manual for Group-Based Designs “Type of comparison group” and “Overall confidence rating on judgment of type of comparison group”).

**A1.5 Other (specify).**
B. Other Design Characteristics (When randomization is used)

B1. Unit of assignment to conditions. Reviewers rate the unit of assignment to intervention and control conditions/groups, including individual, classroom, school, or other.

B2. Type of assignment to conditions. Reviewers select a code that best describes how participants were assigned to control and intervention conditions/groups.

- B2.1 Random after matching, stratification, blocking
- B2.2 Random, simple (includes systematic sampling)
- B2.3 Nonrandom, post hoc matching
- B2.4 Nonrandom, other
- B2.5 Other (specify):
- B2.6 Unknown/insufficient information provided
- B2.7 N/A (randomization not used)

B3. Confidence. Reviewers note their overall confidence in how participants were assigned to conditions/groups.

- B3.1 Very low (little basis)
- B3.2 Low (guess)
- B3.3 Moderate (weak inference)
- B3.4 High (strong inference)
- B3.5 Very high (explicitly stated)
- B3.6 N/A (randomization not used)
- B3.7 Unknown/unable to code

B4. Equivalence. Reviewers note whether equivalence was tested at pretest.

B5. Total sample size. Reviewers code the total sample size at the start of the study.

B6. Intervention sample size. Reviewers code the sample size of the intervention conditions/groups at the start of the study.

B7. Control group sample size. Reviewers code the sample size of the control conditions/groups at the start of the study.

For studies using qualitative research methods, code B8 and B9

B8. Coding. The use of a systematic and clearly articulated approach to coding of qualitative data that can involve inductive and/or deductive analysis, and meets the following criteria:

- B8.1 Coding scheme is linked to study’s theoretical-empirical basis
- B8.2 Procedures for ensuring consistency of coding are used (e.g., dialogue among coders to ensure consistency of meaning and clarify discrepancies; testing the consistency of code application across coders and data).
- B8.3 Progression from abstract concepts to empirical exemplars is clearly articulated.

B9. Interactive process. The processes of data collection and analysis interact such that
preliminary analytical results are subjected to further and more focused data collection for the purposes of confirmation, explanation, and validation. The process of data collection-analysis continues until consistency and depth of understanding are satisfactorily established.

C. Type of Program

The prevention or intervention should be classified according to the Institute of Medicine's (IOM) (1994) classification system. Universal prevention programs are designed to forestall the development of mental health problems and to promoting competence. Often, in universal programs, it is theorized that everyone would benefit from the program. Selective interventions target individuals or subgroups of a population whose risk of developing a problem is significantly higher than average, or individuals who do not respond to universal prevention programs. Indicated interventions target minimal or detectable signs that foreshadow the development of a problem (IOM, 1994; Shinn, Walker, & Stoner, 2002). Intervention or treatments, on the other hand, are provided immediately to treat a serious disorder or problem (Institute of Medicine, 1994).

C1. Universal prevention program
C2. Selective prevention program
C3. Targeted prevention program
C4. Intervention/Treatment
C5. Unknown

D. Stage of Program

Prevention or intervention programs will also be classified according to the developmental stage of the program. Rossi and Freeman (1995) classify programs as model/demonstration, early stage, or established/institutionalized programs. Model/demonstration programs are those programs being implemented for the first time, and are usually small in scale and evaluated using the most rigorous methods possible. Fidelity is a major research question at this stage. Evaluations of intervention and prevention programs occurring in the early stages are likely to use mixed methods (e.g., a combination of qualitative and quantitative criteria). Finally, evaluations of established programs are less likely to investigate fidelity of implementation and are more likely to use nonrandomized designs and alternative-intervention comparison groups.

D1. Model/demonstration programs
D2. Early stage programs
D3. Established/institutionalized programs
D4. Unknown

E. Concurrent or Historical Intervention Exposure

E1. Current exposure. Participant(s) are exposed to another intervention currently (Specify if information is available).
E2. Prior exposure. Participant(s) were previously exposed to other interventions (Specify if information is available).
E3. Unknown. No information is available regarding concurrent or historical intervention exposure.
II. Key Features for Coding Studies and Rating Level of Evidence/Support

The second type of coding includes eight key features of an intervention study that are rated on a four-point scale to specify the level of evidence (i.e., 3=strong evidence/support, 2= promising evidence/support, 1= marginal or weak evidence/support, 0= no evidence/support).

A. Measurement: Issues of Reliability and Validity

Studies will be evaluated regarding the qualities of measures used to establish the effects of an intervention.

A1. Use of outcome measures that produce reliable scores. In single-participant designs, the primary dependent variable is usually repeatedly administered over various phases of the study, with measures of observer agreement (sometimes called observer reliability) determined with an appropriate statistic (see Primavera, Allison, & Alfonso, 1997, for information on observer agreement). In addition, single-participant research studies may involve outcome measures that are administered infrequently, such as checklists or rating scales (e.g., pre and post-intervention). In such cases, conventional psychometric criteria for reliability may be invoked. In the use of traditional measures for research purposes, Wallen and Fraenkel (2001) reported that reliability should be at least .70, and preferably higher. This information must either be reported or referenced in the article. Observable incidence and/or occurrence rates, such as school attendance rates and homework completion rates, and well-known standardized, norm-referenced assessments will be considered reliable measures.

A2. Multi-method. Multiple (i.e., at least two) assessment methods or approaches were used (e.g., observational data, self-reports, teacher ratings) to evaluate Primary outcomes. This rating is not always applicable. For example, observable incidence rates of some behaviors may not require a multi-method assessment.

A3. Multi-source. Measures were obtained from multiple (i.e., at least two) sources (e.g., teachers, parents, self) to evaluate primary outcomes. This rating is not always applicable. For example, observable incidence rates may not require a multi-source assessment.

A4. Validity of measures reported. Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of tests (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Validation involves gathering evidence to provide a sound scientific basis for the interpretation of scores. Table 2 in the group-based manual provides information on different types of validity that can be assessed, as well as some important cultural considerations in assessing validity. Observable incidence and/or occurrence rates, such as school attendance rates and homework completion rates, and well-known standardized, norm-referenced assessments will be considered valid measures.

With .70 as the rule of thumb for minimum reliability, the standards for rating evidence related to the measurement criterion are as follows:

To receive a rating of 3, strong evidence, the investigators must have (a) used instruments that produced a reliability coefficient of .85 or higher; (b) used multiple (i.e., two or more) methods of collecting data; (c) collected data from multiple (i.e., two or more) sources, when appropriate; and (d) presented a case for the validity of the measures used to assess primary outcomes.

To receive a rating of 2, promising evidence, the investigators must have (a) used measures that produced reliable scores (i.e., .85 or higher) for the primary outcomes and (b) used multiple (i.e., two or
more) methods of collecting data, and (c) collected data from multiple (i.e., two or more) sources. A case for validity need not be presented.

To receive a rating of 1, **weak evidence**, the investigators must have used measures that produced somewhat reliable scores (i.e., at least .70) for the primary outcomes under study. In addition, the investigators should have used multiple methods of collecting data or collected data from multiple sources; however, this is not required for a rating of 1. A case for validity need not be presented.

A rating of 0 indicates that the investigators (a) used measures that produced scores with low reliability (i.e., less than .70); (b) did not use multiple methods of collecting data; and (c) did not collect data from multiple sources.

**B. Quality of Baseline**

The baseline phase provides information about the level of a dependent variable before the intervention begins and serves as the standard by which intervention effects are assessed. As such, the baseline phase is a critical component of measurement that has two important functions: (1) it describes the extent of a participant's problems, and (2) it provides a basis for predicting behavior if the intervention was not implemented. According to Hayes et al. (1999), the purpose of repeated assessment during baseline is to establish adequate estimates of level, trend, and stability in the major outcome variables being assessed. Although there is no minimum requirement that is equally appropriate across all target issues/problems, general guidelines are provided below.

**B1. Length of baseline phase(s).** When measures are taken more frequently, there is a greater amount of precision in a time-series. Therefore, it is recommended that at least three data points are needed to begin to estimate the amount of variability in the data, although more measures would be preferred.

**B2. Stability of the baseline phase(s).** Stability refers to fluctuations or variability in a participant’s performance over time. Typically, as variability increases, it becomes more difficult to draw conclusions about intervention effects. The baseline phase(s) are coded as unstable when variability in scores eliminates the detection of treatment effects.

**B3. Overlap of baseline and intervention scores.** To code overlap, extreme scores during baseline should not overlap with most scores during the treatment phase.

**B4. Level of the data.** The level of the problem behavior or measure must be serious enough during the baseline phase(s) to warrant an intervention (i.e., the level of behavior at baseline must be sufficient to allow the intervention to have an impact and show change in the data series).

**B5. Trends in the data (i.e., systematic increases or decreases in a dependent variable).** Baseline data should not demonstrate any slopes or trends in a desired direction of intervention effects.

To receive a rating of 3, **strong evidence**, a study must provide evidence of a high quality baseline for at least 4 of the 5 criteria listed (i.e., length, stability, overlap, level, and trend) for each primary outcome.

To receive a rating of 2, **promising evidence**, a study must provide evidence of a high quality baseline for at least 3 of the 5 criteria listed for each primary outcome.

To receive a rating of 1, **weak evidence**, a study must provide evidence in support of a high quality baseline for at least 2 of the 5 criteria for each primary outcome.
A rating of 0 indicates that one or none of the above criteria were met. Reviewers should provide a rating for quality of baseline: (a) for each participant (when there is more than one participant), and (b) for each baseline phase. Unknown/insufficient data are rated as “no evidence.” In addition, reviewers should compute mean ratings for quality of baseline across all participants, rounding up or down to the nearest whole number (e.g., 2.0 to 2.4 rated as 2, 2.5 to 2.9 rated as 3). These procedures should be followed for each primary outcome under investigation.

C. Measures Support Primary and Secondary Outcomes

To be considered a primary outcome, an outcome measure should represent a valid and appropriate indicator for the type of prevention or intervention program being investigated. Primary outcomes are those that reflect the ultimate goals of the intervention. For example, if an intervention targets reading improvement, than the outcome must measure reading performance. Outcomes restricted to attitudes would not be sufficient to meet the criteria of a primary outcome. Primary outcomes include who changed, what changed, the instrument used to determine change, and the variables expected to change because of the intervention (see group-based manual for an explanation of secondary outcomes).

One important criterion in research is whether the intervention demonstrates a statistically significant effect on the outcome measures following the intervention. Although statistical tests are rarely applied to single-participant designs, in many cases the use of inferential statistics is an option (see Kratochwill & Levin, 1992). When statistical tests are used, our Task Force will consider the guidelines presented by Wilkinson and the Task Force on Statistical Inference of the APA Board of Scientific Affairs (1999). In judging the evidence to support a primary or secondary outcome, our Task Force will use empirical tools to assess statistical significance, such as effect sizes, and, whenever possible, the power of all primary and secondary outcome measures. Outcomes will be considered statistically significant if they reach an alpha level of .05 or less (using familywise error rates when necessary).

To assist reviewers in coding the significance of primary and secondary outcomes, the table found in the coding protocol (“Measures Support Primary and Secondary Outcomes”) should be completed for each participant in the study unless the group is the unit of analysis. Reviewers will also be asked to note any nonsignificant and/or negative outcomes associated with the intervention in a separate table (“Null Findings/Negative Outcomes”).

Primary and secondary outcomes will be coded according to the following categories:

**Outcome.** Specify the outcome being tested, noting primary outcomes before secondary outcomes. All outcomes are listed in alphabetical order.

**Primary versus secondary.** Specify whether the outcome was primary or secondary.

For each outcome, reviewers code the following for each participant:

**For whom the outcome(s) were significant:** Child, teacher, parent/significant adult, ecology (classroom context/peers).

**What changed:** Behavior, attitude, knowledge, other.

**Type of measurement used to determine change:** Self-report, parent report, teacher report, observation, test, other, unknown/no information provided.
Treatment phases used to evaluate the effects of the intervention

Outcome measures used

C1. Evaluating outcomes through visual analysis. In addition to the aforementioned criteria, single-participant designs are most often evaluated on the basis of visual analysis. This analysis refers to the visual examination of the graphical data and the rendering of a judgment about the intervention outcome, and helps evaluate whether a phase change is the result of an intervention and shows improvement in participant behavior. Although there are no exact criteria for visual analysis of data, formalized methods have been developed to assist researchers in the visual analysis process (see Franklin, Gorman, Beasley, & Allison, 1996; Parsonson & Baer, 1992). Evaluating outcomes through visual analysis is coded based on the following categories:

C1.1 Change in levels. Dramatic changes in level between adjacent phases is suggestive of strong intervention effects.

C1.2 Minimal score overlap. Minimal score overlap between the baseline and intervention phases may be suggestive of a strong intervention effect.

C1.3 Change in trend. A change in the trend between adjacent phases that is in the desired direction can also be suggestive of a positive impact of the intervention.

C1.4 Adequate length. At least three data points are needed to estimate the amount of variability in the data, although more measures would be preferred.

C1.5 Stable data. Variability in scores does not eliminate the detection of intervention effects.

C2. Reporting effect sizes. Table 2 outlines three approaches for calculating effect sizes that were developed by Busk and Serlin (1992). Reviewers should record effect size statistics from the original article for each participant, including: type of effect size data, type of data for which the effect size is based, and the effect size statistic. Reviewers should note the approach used to calculate the effect size and confidence ratings in ES computation.

C3. Measures support primary outcomes

To receive a rating of 3, strong evidence, a study must show that measures support the primary outcomes assessed. Using visual analysis, the coder must also find a large change in the level of the behavior during the intervention, in addition to at least three of the following: no-to-minimal overlap between the baseline and intervention phases, a clear trend reversal (if a trend was present), adequate phase length, or stable data. If primary outcomes depend on related outcomes as suggested by the underlying theory, positive changes must be demonstrated for these outcomes as well.

To receive a rating of 2, promising evidence, a study must show that measures support the primary outcomes assessed. Using visual analysis, the coder must also find moderate-to-large changes in the level of behavior during the intervention phase, in addition to at least two of the following: no-to-minimal score overlap, changes in trend (if a trend was present), adequate phase length, or stable data. In addition, if primary outcomes depend on related outcomes as suggested by the underlying theory, positive changes must be demonstrated for these outcomes.
Table 2. Methods to Calculate Single-Participant Effect Sizes

<table>
<thead>
<tr>
<th>Approach</th>
<th>Assumptions</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approach #1</td>
<td>No assumptions concerning population distributional form and equality of intermeasure variances and covariances.</td>
<td>ES = Difference in phase means for baseline and intervention divided by the baseline standard deviation.</td>
</tr>
<tr>
<td>Approach #2</td>
<td>Assumption of equality of variances across baseline and intervention phases.</td>
<td>ES = Pool within phase variances (i.e., within-phase standard deviations are pooled by squaring them to obtain variances), multiplying by the within-phase degrees of freedom, adding the products, and dividing by the overall degrees of freedom.</td>
</tr>
<tr>
<td>Approach #3</td>
<td>Assumptions of a normal distribution and equality of variances and intercorrelations across baseline and intervention phases.</td>
<td>ES = Assuming the phase scores are normally distributed, the within-phase variances are the same, and the within-phase error term follow the noncentral distribution.</td>
</tr>
</tbody>
</table>

To receive a rating of **1**, weak evidence, a study must show that measures support the primary outcomes assessed. Using visual analysis, the coder must also find moderate changes in the level of behavior during the intervention phase, in addition to at least one of the following: no-to-minimal score overlap, changes in trend (if a trend was present), adequate phase length, or stable data.

A rating of **0** would indicate no change in the level of behavior, and that either one or none of the remaining criteria were met. When there is more than one participant, reviewers should provide the rating for primary outcomes for each participant and a mean rating for primary outcomes across all participants, rounding up or down to the nearest whole number (e.g., 2.0 to 2.4 rated as 2, 2.5 to 2.9 rated as 3).

Please note that sometimes single-participant data includes more traditional pre-post measures of statistical significance (used for outcome assessment, not evaluation of social validity). Please note those pre-post measures when they are used.

**C4. Measures Support Secondary Outcomes**

The aforementioned criteria are used to rate secondary outcomes.

**D. Educational/Clinical Significance**

Clinical significance is defined as changes in an individual's behavior that are meaningful and important (Kazdin, 1977). Furthermore, changes in behavior must “resolve or significantly ameliorate the problem for which the client sought treatment” (Kazdin, 1977, p. 427). Clinical significance includes four domains that should be considered when evaluating the quality of evidence related to educational/clinical significance.

**D1. Categorical diagnosis data.** Criteria may be provided regarding inclusion into the study, or changes in diagnostic criteria at baseline, intervention, and follow up phases.
D2. Outcomes assessed via continuous variables. This assessment may include groups or multiple N single-participant designs, and is characterized as the percentage of participants who fall inside the range of dysfunction, the percentage of participants who fall in the normal range, or the percentage of participants showing clinical improvement during each phase of the intervention.

D3. Subjective evaluation. Reviewers should code any information provided regarding the importance of the behavioral change at baseline, intervention, and follow up as evaluated by individuals in direct contact with the participant.

D4. Social comparison. Reviewers should code any information relating to the participant’s behavior at baseline, intervention, and follow up in comparison to normative data such as typical peers.

To receive a rating of 3, strong evidence, a study must provide evidence in support of the clinical significance for at least 3 of the 4 criteria listed (i.e., Categorical Diagnosis Data, Outcomes Assessed Via Continuous Variables, Subjective Evaluation, or Social Comparison) during the intervention or follow up phase for the majority of participants.

To receive a rating of 2, promising evidence, a study must provide evidence in support of the clinical significance for at least 2 of the 4 criteria listed for the majority of participants.

To receive a rating of 1, weak/marginal evidence, a study must provide evidence in support of the clinical significance for at least 1 of the 4 criteria for the majority of participants.

A rating of 0 indicates that none of the above criteria were met.

E. Identifiable Components

Identifiable components focus on the degree to which the authors of a study have identified unique aspects of an intervention program that are related to clinically or statistically significant outcomes. Some interventions include multiple components. In these cases, it must be demonstrated that each component is necessary for the primary outcome to occur. For example, if an intervention designed to prevent teen smoking includes a knowledge component, a peer component, and a parent component, then the study should demonstrate that all three of these components were necessary for the program to effectively prevent teen smoking. As part of this code, reviewers should note the number of program components when that information is available.

To receive a rating of 3, strong evidence, for identifiable components, a study must (1) describe the intervention components in sufficient detail to enable a researcher or practitioner to implement the program as it was implemented in the study, and (2) provide evidence that each intervention component produced significant outcome effects. That is, each intervention component should be linked to a logical measured outcome, and there should be evidence showing that the component distinguished this intervention from another. Referring to a description of such a unique component is acceptable if the description was used to design the study in question (e.g., examining the effect on peer acceptance and social interaction of a social competence intervention with positive peer models versus a social competence intervention with only high-risk students).

To receive a rating of 2, promising evidence, for identifiable components, a study must (1) list and describe the intervention components in sufficient detail to give the reader a comprehensive understanding of the intervention components, and (2) link each intervention component to a logical, significant intervention outcome. A reference to a description of a unique component is acceptable if it is the description used to design the study in question.
To receive a rating of 1, weak evidence, a study must (1) list and briefly describe the intervention components, and (2) provide some indication that the components were viable in producing significant outcomes.

A rating of 0 indicates that the components were not listed or that none of the above criteria were met.

**Additional CRITERIA FOR JUDGING IDENTIFIABLE COMPONENTS (coded descriptively):**

**E5. Documenting essential components.** The procedures for insuring delivery of essential components are described in detail, including the procedures for training facilitators, delivery the intervention, and monitoring the intervention.

**E6. Documenting adaptation.** The procedures for adapting the intervention to fit the ecology (context, culture, participants) are described in detail, including documentation of the impetus (evidence-based occurrence or feedback) for adaptations and the resulting procedural changes. Each adaptation must be clearly linked to data generated by documentation techniques.

**E7. Documenting contextual features.** A detailed description of the context within which the intervention occurred is provided. Context refers to the community in which the intervention takes place (e.g., city, neighborhood); the specific site of implementation (e.g., school); and the specific location (e.g., classroom). Contextual variables include geographic location, population (size, demographics), social-cultural variables (e.g., relevant cultural practices or norms, social issues), and resources relevant to the intervention (e.g., necessary equipment and supplies, expenditures).

**F. Implementation Fidelity**

Reviewers will consider program implementation fidelity/integrity. Although there may be slight variations across implementation, data on program integrity will be critical to determining if the intervention was responsible for the positive outcomes reported by the researchers. Two issues are considered critical here: acceptable adherence and intervention manuals. Reviewers will code the degree to which the authors have confirmed program implementation integrity.

**F1. Acceptable adherence.** The reviewers must be persuaded that the intervention in question was delivered in the study in a manner consistent with the intervention developer’s manual or protocol. Adequate adherence to the intervention developer’s protocol must be shown in all reviewed studies: supporting studies, null studies (i.e., those showing no difference between the candidate intervention and another intervention or baseline control), and refuting studies (i.e., those showing the candidate intervention to be worse than another intervention or control).

**F1.1** When training, ongoing supervision and case consultation are provided by the intervention development team or another expert in the intervention.

**F1.2** When coding of sessions (live, video, or audio) has been done, it must demonstrate that, on average, the majority of the authors’ or the intervention developers’ criteria for adherence were satisfied.

**F1.3** When an intervention is delivered via audio- or videotapes, with or without discussion afterwards, that format demonstrates acceptable adherence.
F2. Manualization. The candidate intervention should be accompanied by a clear description of the procedures used and the studies must be conducted with intervention manuals and/or detailed procedural specification. Manuals may vary in form, depending on the intervention; some manuals may involve a detailed session-by-session account of exact procedures and the sequence in which they are to be used. Others may emphasize broad principles, describe phases of the intervention, and offer only examples of how the broad principles may be implemented. In some cases (e.g., spelling intervention program), the principles and procedures are straightforward enough that they can be described in the Method section of an article, with no need for a separate manual.

In each instance, the test of whether manualization is adequate is (1) whether the intervention procedures are sufficiently well documented that readers can know what intervention procedures were tested, and (2) intervention agents/therapists/consultants/instructors and researchers who wish to replicate the intervention could do so faithfully (recognizing that supplemental training may be required for nuances of the intervention to be understood). This documentation must either be present in the article or the documentation (manual and related materials) and must be available to the interested reader/clinician/researcher.

F3. Adaptation. The study describes the principles and procedures for adapting the intervention to fit varying contexts, cultures and participants while maintaining essential components. Principles of adaptation are evidence-based and derived from repetition of the intervention across multiple contexts, cultures, and samples of participants. The procedural guide advocates repeated evaluation for each application of the intervention.

To receive a rating of 3, strong evidence, the study must demonstrate strong evidence of acceptable adherence. In addition, evidence should be measured through at least two of the following: ongoing supervision/consultation, coding sessions, or audio/video tapes, and use of a manual. Finally, if adaptation occurs to fit varying contexts, there is a description of the procedures for adaptation.

To receive a rating of 2, promising evidence, the study must demonstrate evidence of acceptable adherence. In addition, evidence should be measured through at least one of the above criteria and use of a manual.

To receive a rating of 1, weak evidence, the study must demonstrate evidence of acceptable adherence measured through at least one of the above criteria or use of a manual.

A rating of 0, no evidence, would reflect that nothing was done to ensure implementation fidelity or evidence indicates unacceptable adherence. When there is more than one participant, the rating of fidelity should be an average across all participants, rounding up or down to the nearest whole number when providing an overall rating.

G. Replication

Reviewers code whether the study is a replication of the same intervention, a replication that investigates a similar problem, and whether an independent investigator conducts the replication.

G1. Same intervention. Two versions of an intervention program are considered to be the same if (1) the authors judge the intervention to be essentially the same (declared in the published study, via personal contact, or use of the same manual/procedures) and (2) the intervention duration is equal, the shorter version is at least 75% of the longer version, or a longer intervention is used. If neither or only one of these criteria is satisfied, then the two intervention programs are considered different interventions. Whenever there is a question about whether the intervention is the same (e.g., manual was revised for group vs. individual administration), the reviewers will attempt to retrieve all manuals and related materials from the author(s), review them, and come to a consensus on this intervention issue.
G2. Same target problem. There must be at least two studies meeting all criteria treating the same target issue/problem, and the same age/grade group. First, studies must provide sufficient description of the target issue/problem of the treated group to permit clear identification of the target issue/problem and target age/grade range to which the intervention is relevant. Second, a study is considered a replication only when the intervention is applied to a sample of the same target issue/problem and age/grade range as in the prior study(ies); otherwise, the new study represents a new test with a new target issue/problem and/or developmental group. When an intervention is tested with a different target issue/problem or with a different age/grade group than in previous research, this is considered evidence of a new application of the intervention, not a replication.

The studies will be grouped by age in the following way: infant (birth to 2 years), preschool (2-5 years), elementary (6-11 years), secondary (12-17 years), adult (18-65 years), and geriatric (over 65 years). At transitional points (i.e. 6-11 years), reviewers will present information and recommend whether to consider the participants as the same population, with sub-domain co-chairs ultimately voting on the issue. Some examples of target issue/problem to clarify: if an intervention has two studies supporting its efficacy, one treating depression and the other anxiety, this intervention does not qualify as a replication; if an intervention is supported by two separate studies, one treating conduct problems and oppositionality in children 8-11 years, and the other treating conduct disorder in children 8-11 years, this intervention may be considered replicated, assuming it meets all other criteria. When there is question about whether a study represents a replication with the same or different target issue/problem, the evidence will be presented to a subcommittee and will be voted upon by the committee (see Section 3: Procedures for Decision-Making by Task Force).

G3. Independent evaluation. Reviewers will take into account any relationship between the evaluator/researcher and the intervention program. Evaluations of an intervention conducted by the program developers may show larger effects than those conducted by independent evaluators. Information pertaining to an independent evaluation will be coded and noted in all reviews.

To receive a rating of 3, strong evidence, the study must be a replication of the same intervention and target problem and be implemented by an independent evaluator.

To receive a rating of 2, promising evidence, the study must contain two of the three requirements (i.e., same intervention, same target problem, independent evaluation).

To receive a rating of 1, weak evidence, the study must contain at least one of these requirements.

A rating of 0 means that none of these requirements were met.

H. Site of Implementation

Reviewers will note where the intervention took place. In the rating, preference is given to school settings although it is recognized that interventions may take place in many different sites (e.g., home, university clinic, summer program, outpatient hospital, partial inpatient or day intervention program, inpatient hospital, private practice, mental health center or residential intervention facility). One purpose of recording this information is to indicate whether the intervention took place in (a) a school setting, (b) a clinical setting, where mental health services are routinely delivered independent of the study (e.g., a community mental health center, HMO, inpatient psychiatric unit), or (c) a setting specially arranged for research (e.g., university lab clinic, school classroom).

To receive a rating of 3, strong evidence, the study must have been conducted in a public school or an alternative school, as these are where the majority of children needing intervention are located.

To receive a rating of 2, promising evidence, the study must have been conducted in a private, charter, or university affiliated school setting. These settings all contain special variables, which may not exist in public school
settings and that may contribute to the interventions effectiveness.

To receive a rating of 1, weak evidence, the intervention may have been implemented in a school setting, but could be implemented with little modification (i.e., a video parent training program that requires videotape vignettes be shown to a group of parents, followed by a psychologist facilitated discussion could be done within a school context little modification).

A rating of 0, no evidence, would be an intervention not implemented within a school context, and one that would require major modifications to do so.

II. Follow Up Assessment

Reviewers will code the extent to which follow up assessment was built into the study design. The rating is based on information the authors provide on the timing of outcome assessments, the number of participants included in the follow up assessment, in addition to the consistency of assessment method used.

To receive a rating of 3, strong evidence, the study must have conducted follow up assessments over multiple intervals (e.g., 6 months, 1 year), with all participants that were included in the original sample, using similar measures used to analyze data from primary or secondary outcomes.

To receive a rating of 2, promising evidence, the study must have conducted follow up assessments at least once (e.g., 6 months), with the majority of participants that were included in the original sample, using similar measures used to analyze data from primary or secondary outcomes.

To receive a rating of 1, weak evidence at least once (e.g., 6 months), with some participants from the original sample.

A rating of 0, no evidence, would indicate that no follow up assessment was built into the study.

III. Other Descriptive or Supplemental Criteria to Consider

A. External Validity Indicators

A1. Participant selection (sampling). Participants refer to all persons involved in the intervention, including recipients (e.g., students), implementers (e.g., teachers, psychologists), staff trainers (e.g., researchers, interventionists), decision makers (e.g., principal, research team members), and other key stakeholders who may influence the other participants (e.g., parents, community members, superintendent). Participants are described in sufficient detail to permit other researchers and interventionists to determine both the conditions under which the intervention occurred and the generalizability to intended participants. Reviewers will consider any information reported by the authors regarding the sources from which the sample was drawn, and the means by which clients came to be in the study (e.g., screening at an elementary school, university clinic pool of families, newspaper advertisements, community mental health referral).

A1.1 Inclusion/exclusion criteria specified. Reviewers will consider what participant selection criteria the authors applied to the sample that might limit its representativeness (e.g., whether authors required parent participation, required a two-parent family, or excluded comorbid cases).
A1.2. Inclusion/exclusion criteria similar to school practice. Reviewers will also consider whether inclusion/exclusion criteria were similar to those likely to be used in school practice (code as yes/no). For example, criteria that are similar to school practice might include ruling several physically aggressive students out of a prevention program, and excluding youth from individual psychotherapy who are currently abusing substances. Criteria that are dissimilar from school practice might involve, for example, excluding children who do not have two parents living at home, or eliminating cases with comorbid diagnoses other than the one targeted in intervention.

A1.3 Specified criteria related to concern. The criteria for inclusion into the study should be related to the goal of the intervention. For example, if an intervention is designed to reduce aggressive behavior, then some specified criterion related to the presence of aggressive behavior should be included in the study.

A2. Participant characteristics specified for treatment and control group. Characteristics of the samples used to test interventions will be specified. An objective here is to identify the range of client (e.g., child/family/mediator) characteristics within which an intervention has been shown to have beneficial effects. Each intervention and control sample will be described along the following demographic dimensions:

A2.1 Type of participant. Specify student, parent, teacher, or larger system
A2.2 Grade/age
A2.3 Gender
A2.4 Ethnicity or multiethnic
A2.5 Ethnic identity
A2.6 Race(s)
A2.7 Acculturation
A2.8 Primary language
A2.9 SES
A2.10 Family structure (e.g., single parent, immigrant status)
A2.11 Locale (e.g., urban, rural, suburban, university-affiliated site)
A2.12 Disability (or other special considerations such as English Language Learners)
A2.13 Functional descriptors (e.g., clinical, sub-clinical, "normal" groups; other academic problems present, such as at-risk factors, low reading achievement, etc.)

A3. Details are provided regarding demographic variables (age, gender, ethnicity-race) and any other variables that have been identified as:

A3.1. Having differential relevance to intended outcomes.
A3.2. Relevant to inclusion or exclusion (e.g., level of education, prior experience).

A4. Receptivity/acceptance by target participant population. Please note any consumer satisfaction information, such as reactions from children, parents, teachers, or other program participants as well as reactions to and perceptions of what was gained as a result of the program. Reviewers should also provide a global rating as to whether the participants reported benefiting or not benefiting from the intervention.

A5. Generalization of effects. The extent to which the outcomes generalize across time, settings, and persons. Generalization must be documented through systematic data collection.
A5.1 Generalization over time

A5.1.1 Persistence (sustainability) of outcomes. Evidence is provided regarding the extent to which outcomes persist following the termination of the intervention.

A5.1.2 Procedures for ensuring sustainability. The conditions under which sustainability has been achieved; that is, documentation of efforts (or lack thereof) to ensure maintenance of outcomes (e.g., through booster sessions).

A5.2 Generalization across settings

A5.2.1 Application outside of the intervention context. Evidence is provided regarding the extent to which outcomes are manifested in contexts that are different from the intervention context. For example, evidence of the capacity for the participants to make use of newly acquired skills not only within the intervention context (e.g., resolving conflicts with peers in the classroom) but also in other settings (e.g., on the playground) or applied to other types of problems (e.g., solving conflicts with adults).

A5.2.2 Procedures for ensuring application. The conditions under which application outside of the intervention context was achieved; that is, documentation of efforts (or lack thereof) to ensure application to other settings, problems, etc. (e.g., through simulated practice in wide scale application; booster sessions in other contexts).

A5.2.3 Institutionalization. The extent to which impact on implementers or context is sustained, documented through systematic follow-up. For example, the extent to which teachers continue the application of a classroom-based intervention after the end of the formal intervention program; or the extent to which the school system continues discipline practices established by the intervention program. The conditions under which sustainability has been achieved; that is, documentation of efforts (or lack thereof) to ensure sustained impact (e.g., through booster sessions or follow-up consultation).

A5.3 Generalization across persons

Evidence is provided regarding the extent to which outcomes are manifested with participants who are different than the original group of participants for which the intervention was evaluated. An underlying assumption is that each implementation of an intervention in a different context or with a different group constitutes a “different” intervention, given the cultural-contextual nature of interventions.

B. Length of Intervention

Reviewers will code the length of the intervention (i.e., how long the intervention was in place) in weeks, months, or years.

C. Intensity/Dosage of Intervention

Intensity refers to the length of the intervention sessions and the frequency of the sessions. Reviewers will code both length and frequency of intervention sessions.
D. Dosage/Response

Reviewers will note whether there is evidence that length or intensity of intervention was associated with stronger outcomes or better performance on outcome measures.

E. Program Implementer

Reviewers will code who conducted the intervention. Possible choices include: research staff, school specialty staff (e.g., counselors, psychologists, social workers, special educators, etc.), teachers, educational assistants, parents, college students, peers, and others. If research staff were only involved in training intervention implementers, then do not include them as program implementers, unless the intervention is the training. This support will be documented under Training and Support Resources.

F. Characteristics of the Intervener

Reviewers will note how similar the target participants were to the intervener on several characteristics (e.g., race, gender, SES).

G. Intervention Style or Orientation

Reviewers will note the theoretical underpinnings of the intervention program, such as behavioral, cognitive-behavioral, experiential, humanistic/interpersonal, psychodynamic/insight oriented, or combination/other.

H. Cost Analysis Data

Cost analysis data provided (coded yes/no). Whenever authors report data regarding the cost to implement the program, and/or cost-to-benefit analyses, reviewers will report this information.

I. Training and Support Resources

What supports were provided if school or other typical staff implemented the intervention in the study?

I1. Simple orientation given to change agents

I2. Training workshops given (indicate # of workshops, average length of training, and who conducted the training)

I3. On-going technical support provided once intervention began, such as phone or face-to-face consultation, classroom observations, coaching or trouble-shooting after program has begun

I4. Program materials must be obtained to conduct intervention

I5. Special facilities (extra space) must be obtained

I6. Other
J. Feasibility

Reviewers will rate the level of difficulty in training intervention agents (i.e., high, moderate, or low) and report and rate the cost involved in training intervention agents (i.e., high, moderate, low) when this information is available.
Coding Protocol: Single-Participant Design

Domain: □ School- and community-based intervention programs for social and behavioral problems
□ Academic intervention programs
□ Family and parent intervention programs
□ School-wide and classroom-based programs
□ Comprehensive and coordinated school health services

Name of Coder(s): ______________________________  Date: ____________________________  M / D / Y

Full Study Reference in APA format: __________________________________________________________
________________________________________________________________________________________
________________________________________________________________________________________

Intervention Name (description from study): ____________________________________________________
________________________________________________________________________________________
________________________________________________________________________________________

Study ID Number (Unique Identifier): __________________________________________________________

Type of Publication:  (Check one)

□ Book/Monograph
□ Journal article
□ Book chapter
□ Other (specify):
I. General Characteristics

A. General Design Characteristics (Classify studies according to the type of design)


A1.1  □  Within-series design (select A1.1.1 or A1.1.2)
  A1.1.1  □  Simple phase change
  A1.1.2  □  Complex phase change

A1.2  □  Between-series design (select A1.2.1 or A1.2.2)
  A1.2.1  □  Comparing two interventions
  A1.2.2  □  Comparing interventions with no interventions

A1.3  □  Combined-series design (select A1.3.1., A1.3.2, A1.3.3, or A1.3.4)
  A1.3.1  □  Multiple baseline across participants
  A1.3.2  □  Multiple baseline across behaviors
  A1.3.3  □  Multiple baseline across settings
  A1.3.4  □  Multiple probe design

A1.4  □  Mixed design (select A1.4.1 or A1.4.2)
  A1.4.1  □  Combined single-participant and group design (see group manual),
  A1.4.2  □  Combined single-participant design (if combined single-participant design, check A1.4.2.1, A1.4.2.2, or A1.4.2.3)

  A1.4.2.1  □  Within-series design (select i or ii)
    i.  □  Simple phase change
    ii.  □  Complex phase change

  A1.4.2.2  □  Between-series design (select i or ii)
    i.  □  Comparing two interventions
    ii.  □  Comparing interventions with no interventions

  A1.4.2.3  □  Combined-series design (select i, ii, iii, or iv)
    i.  □  Multiple baseline across participants
    ii.  □  Multiple baseline across behaviors
    iii.  □  Multiple baseline across settings
    iv.  □  Multiple probe design

A1.5  □  Other (specify):
B. Other Design Characteristics (when randomization is used)

B1. Unit of assignment to conditions/groups (select one of the following)

- B1.1 Individual
- B1.2 Classroom
- B1.3 School
- B1.4 Other (specify):
- B1.5 N/A (randomization not used)

B2. Type of assignment to conditions/groups (select one of the following)

- B2.1 Random after matching, stratification, blocking
- B2.2 Random, simple (includes systematic sampling)
- B2.3 Nonrandom, post hoc matching
- B2.4 Nonrandom, other
- B2.5 Other (specify):
- B2.6 Unknown/insufficient information provided
- B2.7 N/A (randomization not used)

B3. Overall confidence of judgment on how participants were assigned to conditions/groups (select one of the following)

- B3.1 Very low (little basis)
- B3.2 Low (guess)
- B3.3 Moderate (weak inference)
- B3.4 High (strong inference)
- B3.5 Very high (explicitly stated)
- B3.6 N/A (randomization not used)
- B3.7 Unknown/unable to code

B4. Equivalence of conditions/groups tested at pretest (select one of the following)

- B4.1 Yes
- B4.2 No
- B4.3 Unknown/insufficient information provided
- B4.4 N/A (randomization not used)

B5. Total size of sample (start of the study): ________________

B6. Intervention sample size ________________ N/A (randomization not used)

B7. Control sample size ________________ N/A (randomization not used)

For studies using qualitative research methods, code B8 and B9

B8. Coding

- B8.1 Coding scheme linked to study’s theoretical-empirical basis (select one)  yes  no
B8.2 Procedures for ensuring consistency of coding are used (select one)  □ yes  □ no

Describe procedures: ________________________________________________________________

B8.3 Progression from abstract concepts to empirical exemplars is clearly articulated (select one)
□ yes  □ no

B9. Interactive process followed (select one)  □ yes  □ no

Describe process: ________________________________________________________________

C. Type of Program (select one)

C1. □ Universal prevention program
C2. □ Selective prevention program
C3. □ Targeted prevention program
C4. □ Intervention/Treatment
C5. □ Unknown

D. Stage of the Program (select one)

D1. □ Model/demonstration programs
D2. □ Early stage programs
D3. □ Established/institutionalized programs
D4. □ Unknown

E. Concurrent or Historical Intervention Exposure (select one)

E1. □ Current exposure
E2. □ Prior exposure
E3. □ Unknown
## II. Key Features for Coding Studies and Rating Level of Evidence

(3=Strong Evidence    2=Promising Evidence    1=Weak Evidence    0=No Evidence)

### A. Measurement: Issues of Reliability and Validity (answer A1. through A4.)

A1. Use of outcome measures that produce reliable scores (select one of the following)

- A1.1 Yes
- A1.2 No
- A1.3 Unknown/unable to code

A2. Multi-method (select one of the following)

- A2.1 Yes
- A2.2 No
- A2.3 N/A
- A2.4 Unknown/unable to code

A3. Multi-source (select one of the following)

- A3.1 Yes
- A3.2 No
- A3.3 N/A
- A3.4 Unknown/unable to code

A4. Validity of measures reported (select one of the following)

- A4.1 Yes
- A4.2 No
- A4.3 Unknown/unable to code

**Rating for Measurement** (select 0, 1, 2, or 3): 3 2 1 0

### B. Quality of Baseline.

Rate quality of baseline: (a) for each participant (when there is more than one participant), and (b) for each phase (when the study includes more than one phase). These procedures should be followed for each primary outcome under investigation.

**Participant 1** (answer B1. through B5.)

B1. Length: At least 3 data points during baseline (select one of the following)

- B1.1 Yes
- B1.2 No
- B1.3 Unknown/insufficient information provided
B2. Stability: Variability in scores does not eliminate the detection of treatment effects (select one of the following)

B2.1  Yes
B2.2  No
B2.3  Unknown/insufficient information provided

B3. Overlap: Extreme scores during baseline do not overlap with most scores during intervention phase (select one of the following)

B3.1  Yes
B3.2  No
B3.3  Unknown/insufficient information provided

B4. Level: Behavior is serious enough during baseline to warrant an intervention (select one of the following)

B4.1  Yes
B4.2  No
B4.3  Unknown/insufficient information provided

B5. Trend: Behavior is not systematically increasing or decreasing in the desired direction of intervention effects during baseline.

B5.1  Yes
B5.2  No
B5.3  Unknown/insufficient information provided

Rating of quality of baseline for participant 1: (select 0, 1, 2, or 3):  □ 3  □ 2  □ 1  □ 0

Participant 2 (answer B1. through B5.)

B1. Length: At least 3 data points during baseline (select one of the following)

B1.1  Yes
B1.2  No
B1.3  Unknown/insufficient information provided

B2. Stability: Variability in scores does not eliminate the detection of treatment effects (select one of the following)

B2.1  Yes
B2.2  No
B2.3  Unknown/insufficient information provided

B3. Overlap: Extreme scores during baseline do not overlap with most scores during intervention phase (select one of the following)

B3.1  Yes
B3.2  No
B3.3  □ Unknown/insufficient information provided

B4. Level: Behavior is serious enough during baseline to warrant an intervention (select one of the following)

B4.1  □ Yes
B4.2  □ No
B4.3  □ Unknown/insufficient information provided

B5. Trend: Behavior is not systematically increasing or decreasing in the desired direction of intervention effects during baseline.

B5.1  □ Yes
B5.2  □ No
B5.3  □ Unknown/insufficient information provided

Rating of quality of baseline for participant 2: (select 0, 1, 2, or 3):  □ 3  □ 2  □ 1  □ 0

Participant 3 (answer B1. through B5.)

B1. Length: At least 3 data points during baseline (select one of the following)

B1.1  □ Yes
B1.2  □ No
B1.3  □ Unknown/insufficient information provided

B2. Stability: Variability in scores does not eliminate the detection of treatment effects (select one of the following)

B2.1  □ Yes
B2.2  □ No
B2.3  □ Unknown/insufficient information provided

B3. Overlap: Extreme scores during baseline do not overlap with most scores during intervention phase (select one of the following)

B3.1  □ Yes
B3.2  □ No
B3.3  □ Unknown/insufficient information provided

B4. Level: Behavior is serious enough during baseline to warrant an intervention (select one of the following)

B4.1  □ Yes
B4.2  □ No
B4.3  □ Unknown/insufficient information provided

B5. Trend: Behavior is not systematically increasing or decreasing in the desired direction of intervention effects during baseline.

B5.1  □ Yes
B5.2  □ No
B5.3 □ Unknown/insufficient information provided

Rating of quality of baseline for participant 3: (select 0, 1, 2, or 3): □ 3 □ 2 □ 1 □ 0

Participant 4 (answer B1. through B5.)

B1. Length: At least 3 data points during baseline (select one of the following)
   B1.1 □ Yes
   B1.2 □ No
   B1.3 □ Unknown/insufficient information provided

B2. Stability: Variability in scores does not eliminate the detection of treatment effects (select one of the following)
   B2.1 □ Yes
   B2.2 □ No
   B2.3 □ Unknown/insufficient information provided

B3. Overlap: Extreme scores during baseline do not overlap with most scores during intervention phase (select one of the following)
   B3.1 □ Yes
   B3.2 □ No
   B3.3 □ Unknown/insufficient information provided

B4. Level: Behavior is serious enough during baseline to warrant an intervention (select one of the following)
   B4.1 □ Yes
   B4.2 □ No
   B4.3 □ Unknown/insufficient information provided

B5. Trend: Behavior is not systematically increasing or decreasing in the desired direction of intervention effects during baseline.
   B5.1 □ Yes
   B5.2 □ No
   B5.3 □ Unknown/insufficient information provided

Rating of quality of baseline for participant 4: (select 0, 1, 2, or 3): □ 3 □ 2 □ 1 □ 0

Average Quality of Baseline Rating Across Participants:

\[ \Sigma \text{ of } X = \frac{\Sigma \text{ of } X}{N} \]

\[ X = \text{individual quality of baseline ratings for each participant} \]

\[ N = \text{number of participants in the study} \]
Overall Rating for Quality of Baseline: (select 0, 1, 2, or 3): □ 3 □ 2 □ 1 □ 0

(Round up or down to the nearest whole number when providing a mean rating for the study. For example, 2.0 to 2.4 rated as 2; 2.5 to 2.9 rated as 3).
### C. Measures Support Primary and Secondary Outcomes

(list primary outcomes first in alphabetical order for each participant followed by secondary outcomes. Use extra tables if necessary).

**Outcome #1: ________________________________

<table>
<thead>
<tr>
<th>Primary vs. Secondary</th>
<th>Who changed</th>
<th>What changed</th>
<th>Measurement used to determine change</th>
<th>Describe outcome variables</th>
<th>Treatment phases</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Primary</td>
<td>Participant 1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Unknown</td>
<td>Teacher</td>
<td>Attitude</td>
<td>Parent Report</td>
<td>Teacher Report</td>
<td></td>
</tr>
<tr>
<td>□ Unknown</td>
<td>Parent/sign. adult</td>
<td>Knowledge</td>
<td>Observation</td>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>□ Unknown</td>
<td>Ecology</td>
<td>Other</td>
<td>Other</td>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>□ Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>□ Primary</td>
<td>Participant 2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Unknown</td>
<td>Teacher</td>
<td>Attitude</td>
<td>Parent Report</td>
<td>Teacher Report</td>
<td></td>
</tr>
<tr>
<td>□ Unknown</td>
<td>Parent/sign. adult</td>
<td>Knowledge</td>
<td>Observation</td>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>□ Unknown</td>
<td>Ecology</td>
<td>Other</td>
<td>Other</td>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>□ Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>□ Primary</td>
<td>Participant 3:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Unknown</td>
<td>Teacher</td>
<td>Attitude</td>
<td>Parent Report</td>
<td>Teacher Report</td>
<td></td>
</tr>
<tr>
<td>□ Unknown</td>
<td>Parent/sign. adult</td>
<td>Knowledge</td>
<td>Observation</td>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>□ Unknown</td>
<td>Ecology</td>
<td>Other</td>
<td>Other</td>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>□ Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>□ Primary</td>
<td>Participant 4:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Unknown</td>
<td>Teacher</td>
<td>Attitude</td>
<td>Parent Report</td>
<td>Teacher Report</td>
<td></td>
</tr>
<tr>
<td>□ Unknown</td>
<td>Parent/sign. adult</td>
<td>Knowledge</td>
<td>Observation</td>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>□ Unknown</td>
<td>Ecology</td>
<td>Other</td>
<td>Other</td>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>□ Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td></td>
</tr>
</tbody>
</table>
Outcome #2: ____________________________________________

<table>
<thead>
<tr>
<th>Primary vs. Secondary</th>
<th>Who changed</th>
<th>What changed</th>
<th>Measurement used to determine change</th>
<th>Describe outcome variables</th>
<th>Treatment phases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>Participant 1: Child</td>
<td>Behavior</td>
<td>Self Report</td>
<td>Parent Report</td>
<td>Teacher Report</td>
</tr>
<tr>
<td>Secondary</td>
<td>Teacher</td>
<td>Attitude</td>
<td>Parent Report</td>
<td>Teacher Report</td>
<td>Observation</td>
</tr>
<tr>
<td>Unknown</td>
<td>Parent/sign. adult</td>
<td>Knowledge</td>
<td>Observation</td>
<td>Test</td>
<td>Other</td>
</tr>
<tr>
<td>Unknown</td>
<td>Ecology</td>
<td>Other</td>
<td>Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Primary               | Participant 2: Child | Behavior | Self Report | Parent Report | Teacher Report | Observation | Test | Other | Unknown |
| Secondary             | Teacher | Attitude | Parent Report | Teacher Report | Observation | Test | Other | Unknown |
| Unknown               | Parent/sign. adult | Knowledge | Observation | Test | Other | Unknown |
| Unknown               | Ecology | Other | Unknown | |
| Unknown               | Unknown | Unknown | | |

| Secondary             | Teacher | Attitude | Parent Report | Teacher Report | Observation | Test | Other | Unknown |
| Unknown               | Parent/sign. adult | Knowledge | Observation | Test | Other | Unknown |
| Unknown               | Ecology | Other | Unknown | |
| Unknown               | Unknown | Unknown | | |

| Primary               | Participant 4: Child | Behavior | Self Report | Parent Report | Teacher Report | Observation | Test | Other | Unknown |
| Secondary             | Teacher | Attitude | Parent Report | Teacher Report | Observation | Test | Other | Unknown |
| Unknown               | Parent/sign. adult | Knowledge | Observation | Test | Other | Unknown |
| Unknown               | Ecology | Other | Unknown | |
| Unknown               | Unknown | Unknown | | |
### Outcome #3: ____________________________________________

<table>
<thead>
<tr>
<th>Primary vs. Secondary</th>
<th>Who changed</th>
<th>What changed</th>
<th>Measurement used to determine change</th>
<th>Describe outcome variables</th>
<th>Treatment phases</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Primary</td>
<td>Participant 1:</td>
<td>Behavior</td>
<td>□ Self Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Secondary</td>
<td>Child</td>
<td>Attitude</td>
<td>□ Parent Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Unknown</td>
<td>Teacher</td>
<td>Knowledge</td>
<td>□ Teacher Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parent/sign. adult</td>
<td>Other</td>
<td>□ Observation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td>Unknown</td>
<td>□ Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
<td>□ Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>□ Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Primary</td>
<td>Participant 2:</td>
<td>Behavior</td>
<td>□ Self Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Secondary</td>
<td>Child</td>
<td>Attitude</td>
<td>□ Parent Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Unknown</td>
<td>Teacher</td>
<td>Knowledge</td>
<td>□ Teacher Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parent/sign. adult</td>
<td>Other</td>
<td>□ Observation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td>Unknown</td>
<td>□ Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
<td>□ Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>□ Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Primary</td>
<td>Participant 3:</td>
<td>Behavior</td>
<td>□ Self Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Secondary</td>
<td>Child</td>
<td>Attitude</td>
<td>□ Parent Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Unknown</td>
<td>Teacher</td>
<td>Knowledge</td>
<td>□ Teacher Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parent/sign. adult</td>
<td>Other</td>
<td>□ Observation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td>Unknown</td>
<td>□ Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
<td>□ Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>□ Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Primary</td>
<td>Participant 4:</td>
<td>Behavior</td>
<td>□ Self Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Secondary</td>
<td>Child</td>
<td>Attitude</td>
<td>□ Parent Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Unknown</td>
<td>Teacher</td>
<td>Knowledge</td>
<td>□ Teacher Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parent/sign. adult</td>
<td>Other</td>
<td>□ Observation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td>Unknown</td>
<td>□ Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
<td>□ Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>□ Unknown</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Null Findings/Negative Outcomes Associated with the Intervention (listed alphabetically by outcome)

Outcome #1: ____________________________________________

<table>
<thead>
<tr>
<th>Primary vs. Secondary Outcome</th>
<th>Who changed</th>
<th>What changed</th>
<th>Measurement used to determine change</th>
<th>Describe outcome variables</th>
<th>Treatment phases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>Participant 1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Teacher</td>
<td>Attitude</td>
<td>Teacher Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parent/sign. adult</td>
<td>Knowledge</td>
<td>Observation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td>Other</td>
<td>Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>Unkn</td>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>Participant 2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Teacher</td>
<td>Attitude</td>
<td>Teacher Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parent/sign. adult</td>
<td>Knowledge</td>
<td>Observation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td>Other</td>
<td>Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>Unkn</td>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>Participant 3:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Teacher</td>
<td>Attitude</td>
<td>Teacher Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parent/sign. adult</td>
<td>Knowledge</td>
<td>Observation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td>Other</td>
<td>Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>Unkn</td>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>Participant 4:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Teacher</td>
<td>Attitude</td>
<td>Teacher Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parent/sign. adult</td>
<td>Knowledge</td>
<td>Observation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td>Other</td>
<td>Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>Unkn</td>
<td>Other</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
C1. **Visual Analysis** (complete the following tables for each primary and secondary outcome)

**Outcome #1:** ____________________________________________

<table>
<thead>
<tr>
<th>Participant</th>
<th>Primary vs. Secondary Outcome</th>
<th>Visual Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participant 1</strong></td>
<td></td>
<td>(Answer C1.1 through C1.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1 Change in levels (select one of the following)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1.1 large</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1.2 moderate-to-large</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1.3 moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1.4 no change</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1.5 unknown/insufficient information provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.2 No-to-minimal score overlap (select one of the following)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.2.1 yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.2.2 no</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.2.3 unknown/insufficient information provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.3 Change in trend (select one of the following)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.3.1 yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.3.2 no</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.3.3 N/A (no trend present)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.3.4 unknown/insufficient information provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.4 Adequate length (select one of the following)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.4.1 yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.4.2 no</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.4.3 unknown/insufficient information provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.5 Stable data (select one of the following)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.5.1 yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.5.2 no</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.5.3 unknown/insufficient information provided</td>
</tr>
<tr>
<td></td>
<td>Rating for Visual Analysis for Participant 1: (select 0, 1, 2, or 3)</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participant 2</th>
<th></th>
<th>(Answer C1.1 through C1.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>C1.1 Change in levels (select one of the following)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1.1 large</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1.2 moderate-to-large</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1.3 moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1.4 no change</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1.5 unknown/insufficient information provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.2 No-to-minimal score overlap (select one of the following)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.2.1 yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.2.2 no</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.2.3 unknown/insufficient information provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.3 Change in trend (select one of the following)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.3.1 yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.3.2 no</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.3.3 N/A (no trend present)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.3.4 unknown/insufficient information provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.4 Adequate length (select one of the following)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.4.1 yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.4.2 no</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.4.3 unknown/insufficient information provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.5 Stable data (select one of the following)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.5.1 yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.5.2 no</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.5.3 unknown/insufficient information provided</td>
</tr>
<tr>
<td></td>
<td>Rating for Visual Analysis for Participant 2: (select 0, 1, 2, or 3)</td>
<td>0</td>
</tr>
<tr>
<td>Participant</td>
<td>Primary vs. Secondary Outcome</td>
<td>Visual Analysis</td>
</tr>
<tr>
<td>-------------</td>
<td>------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Participant 3</td>
<td>□ Primary</td>
<td>(Answer C1.1 through C1.5)</td>
</tr>
<tr>
<td></td>
<td>□ Secondary</td>
<td></td>
</tr>
<tr>
<td></td>
<td>□ Unknown</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1 Change in levels (select one of the following)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1.1 □ large</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1.2 □ moderate-to-large</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1.3 □ moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1.4 □ no change</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1.5 □ unknown/insufficient information provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.2 No-to-minimal score overlap (select one of the following)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.2.1 □ yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.2.2 □ no</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.2.3 □ unknown/insufficient information provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.3 Change in trend (select one of the following)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.3.1 □ yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.3.2 □ no</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.3.3 □ N/A (no trend present)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.3.4 □ unknown/insufficient information provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.4 Adequate length (select one of the following)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.4.1 □ yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.4.2 □ no</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.4.3 □ unknown/insufficient information provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.5 Stable data (select one of the following)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.5.1 □ yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.5.2 □ no</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.5.3 □ unknown/insufficient information provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rating for Visual Analysis for Participant 3: (select 0, 1, 2, or 3) □0 □1 □2 □3</td>
</tr>
<tr>
<td>Participant 4</td>
<td>□ Primary</td>
<td>(Answer C1.1 through C1.5)</td>
</tr>
<tr>
<td></td>
<td>□ Secondary</td>
<td></td>
</tr>
<tr>
<td></td>
<td>□ Unknown</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1 Change in levels (select one of the following)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1.1 □ large</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1.2 □ moderate-to-large</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1.3 □ moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1.4 □ no change</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.1.5 □ unknown/insufficient information provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.2 No-to-minimal score overlap (select one of the following)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.2.1 □ yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.2.2 □ no</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.2.3 □ unknown/insufficient information provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.3 Change in trend (select one of the following)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.3.1 □ yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.3.2 □ no</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.3.3 □ N/A (no trend present)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.3.4 □ unknown/insufficient information provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.4 Adequate length (select one of the following)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.4.1 □ yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.4.2 □ no</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.4.3 □ unknown/insufficient information provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.5 Stable data (select one of the following)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.5.1 □ yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.5.2 □ no</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C1.5.3 □ unknown/insufficient information provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rating for Visual Analysis for Participant 4: (select 0, 1, 2, or 3) □0 □1 □2 □3</td>
</tr>
</tbody>
</table>
C2. Reporting Effect Sizes (complete the following table for each primary and secondary outcome)

<table>
<thead>
<tr>
<th>Participants</th>
<th>Primary vs. Secondary Outcome</th>
<th>Type of ES Data</th>
<th>Type of Data ES Based On</th>
<th>Effect Size</th>
<th>Confidence Rating in ES Computation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Approach Used</td>
<td></td>
</tr>
<tr>
<td>Participant 1:</td>
<td>Primary</td>
<td>Secondary</td>
<td>Unknown</td>
<td>Baseline vs. intervention (e.g., $A_1 + A_2$ vs. $B_1 + B_2$)</td>
<td>Means and SDs</td>
</tr>
<tr>
<td></td>
<td>Baseline vs. follow up</td>
<td>Unknown</td>
<td></td>
<td>Approach 1</td>
<td></td>
</tr>
<tr>
<td>Participant 2:</td>
<td>Primary</td>
<td>Secondary</td>
<td>Unknown</td>
<td>Baseline vs. intervention (e.g., $A_1 + A_2$ vs. $B_1 + B_2$)</td>
<td>Means and SDs</td>
</tr>
<tr>
<td></td>
<td>Baseline vs. follow up</td>
<td>Unknown</td>
<td></td>
<td>Approach 1</td>
<td></td>
</tr>
<tr>
<td>Participant 3:</td>
<td>Primary</td>
<td>Secondary</td>
<td>Unknown</td>
<td>Baseline vs. intervention (e.g., $A_1 + A_2$ vs. $B_1 + B_2$)</td>
<td>Means and SDs</td>
</tr>
<tr>
<td></td>
<td>Baseline vs. follow up</td>
<td>Unknown</td>
<td></td>
<td>Approach 1</td>
<td></td>
</tr>
<tr>
<td>Participant 4:</td>
<td>Primary</td>
<td>Secondary</td>
<td>Unknown</td>
<td>Baseline vs. intervention (e.g., $A_1 + A_2$ vs. $B_1 + B_2$)</td>
<td>Means and SDs</td>
</tr>
<tr>
<td></td>
<td>Baseline vs. follow up</td>
<td>Unknown</td>
<td></td>
<td>Approach 1</td>
<td></td>
</tr>
</tbody>
</table>
C3. Measures Support Primary Outcomes

Average Rating for Measures Support Primary Outcomes Across Participants Using Score from Visual Analysis:

\[
\sum \frac{X}{N} =
\]

\(X\) = individual primary outcome ratings for each participant
\(N\) = number of participants in the study

Overall Rating for Measures Support Primary Outcomes: (select 0, 1, 2, or 3): 
\(\square\) 3 \(\square\) 2 \(\square\) 1 \(\square\) 0

(Round up or down to the nearest whole number when providing a mean rating for the study. For example, 2.0 to 2.4 rated as 2; 2.5 to 2.9 rated as 3).

C4. Measures Support Secondary Outcomes

Average Rating for Measures Support Secondary Outcomes Across Participants Using Score from Visual Analysis:

\[
\sum \frac{X}{N} =
\]

\(X\) = individual primary outcome ratings for each participant
\(N\) = number of participants in the study

Overall Rating for Measures Support Secondary Outcomes: (select 0, 1, 2, or 3): 
\(\square\) 3 \(\square\) 2 \(\square\) 1 \(\square\) 0

(Round up or down to the nearest whole number when providing a mean rating for the study. For example, 2.0 to 2.4 rated as 2; 2.5 to 2.9 rated as 3).
D. Educational/Clinical Significance

<table>
<thead>
<tr>
<th>Outcome Variables:</th>
<th>Baseline Phase</th>
<th>Intervention Phase</th>
<th>Follow Up Phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1. Categorical Diagnosis Data</td>
<td>Diagnostic information regarding inclusion into the study presented:</td>
<td>Positive change in diagnostic criteria from baseline to intervention</td>
<td>Positive change in diagnostic criteria from intervention to follow up:</td>
</tr>
<tr>
<td></td>
<td>□ Yes □ No □ Unknown</td>
<td>□ Yes □ No □ Unknown</td>
<td>□ Yes □ No □ Unknown</td>
</tr>
<tr>
<td>D2. Outcome Assessed via continuous Variables</td>
<td></td>
<td>Positive change in number of participants showing clinical improvement from baseline to intervention:</td>
<td>Positive change in number of participants showing clinical improvement from intervention to follow up:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Yes □ No □ Unknown</td>
<td>□ Yes □ No □ Unknown</td>
</tr>
<tr>
<td>D3. Subjective Evaluation: The importance of behavior change is evaluated by individuals in direct contact with the participant.</td>
<td>Importance of behavior change is evaluated:</td>
<td>Importance of behavior change from baseline to intervention is evaluated positively by individuals in direct contact with the participant:</td>
<td>Importance of behavior change from intervention to follow up is evaluated positively by individuals in direct contact with the participant:</td>
</tr>
<tr>
<td></td>
<td>□ Yes □ No □ Unknown</td>
<td>□ Yes □ No □ Unknown</td>
<td>□ Yes □ No □ Unknown</td>
</tr>
<tr>
<td>D4. Social Comparison: Behavior of participant at baseline, intervention, and follow up phase is compared to normative data (e.g., a typical peer).</td>
<td>Participant's behavior is compared to normative data</td>
<td>Participant’s behavior has improved from baseline to treatment when compared to normative data:</td>
<td>Participant’s behavior has improved from intervention to follow up when compared to normative data:</td>
</tr>
<tr>
<td></td>
<td>□ Yes □ No □ Unknown</td>
<td>□ Yes □ No □ Unknown</td>
<td>□ Yes □ No □ Unknown</td>
</tr>
</tbody>
</table>

Rating for Educational/Clinical Significance (select 0, 1, 2, or 3): □ 3 □ 2 □ 1 □ 0
E. Identifiable Components (answer E1 through E7)

E1. Evidence for primary outcomes (rate from previous code): 3 2 1 0

E2. Design allows for analysis of identifiable components (select one) yes no

E3. Total number of components: N

E4. Number of components linked to primary outcomes: N

Additional criteria to code descriptively:

E5. Clear documentation of essential components (select one) yes no

E6. Procedures for adapting the intervention are described in detail (select one) yes no

E7. Contextual features of the intervention are documented (select one) yes no

Rating for Identifiable Components (select 0, 1, 2, or 3): 3 2 1 0

F. Implementation Fidelity

Participant 1:

F1. Evidence of Acceptable Adherence (answer F1.1 through F1.3)

F1.1 Ongoing supervision/consultation
F1.2 Coding intervention sessions/lessons or procedures
F1.3 Audio/video tape implementation (select F1.3.1 or F1.3.2):

F1.3.1 Entire intervention
F1.3.2 Part of intervention

F2. Manualization (select all that apply)

F2.1 Written material involving a detailed account of the exact procedures and the sequence in which they are to be used

F2.2 Formal training session that includes a detailed account of the exact procedures and the sequence in which they are to be used

F2.3 Written material involving an overview of broad principles and a description of the intervention phases

F2.4 Formal or informal training session involving an overview of broad principles and a description of the intervention phases

F3. Adaptation procedures are specified (select one) yes no unknown

Rating for Implementation Fidelity for Participant 1 (select 0, 1, 2, or 3): 3 2 1 0
Participant 2:

F1. Evidence of Acceptable Adherence (answer F1.1 through F1.3)

F1.1 ☐ Ongoing supervision/consultation
F1.2 ☐ Coding intervention sessions/lessons or procedures
F1.3 ☐ Audio/video tape implementation (select F1.3.1 or F1.3.2):

F1.3.1 ☐ Entire intervention
F1.3.2 ☐ Part of intervention

F2. Manualization (select all that apply)

F2.1 ☐ Written material involving a detailed account of the exact procedures and the sequence in which they are to be used
F2.2 ☐ Formal training session that includes a detailed account of the exact procedures and the sequence in which they are to be used
F2.3 ☐ Written material involving an overview of broad principles and a description of the intervention phases
F2.4 ☐ Formal or informal training session involving an overview of broad principles and a description of the intervention phases

F3. Adaptation procedures are specified (select one) ☐ yes ☐ no ☐ unknown

Rating for Implementation Fidelity for Participant 2 (select 0, 1, 2, or 3): ☐ 3 ☐ 2 ☐ 1 ☐ 0

Participant 3:

F1. Evidence of Acceptable Adherence (answer F1.1 through F1.3)

F1.1 ☐ Ongoing supervision/consultation
F1.2 ☐ Coding intervention sessions/lessons or procedures
F1.3 ☐ Audio/video tape implementation (select F1.3.1 or F1.3.2):

F1.3.1 ☐ Entire intervention
F1.3.2 ☐ Part of intervention

F2. Manualization (select all that apply)

F2.1 ☐ Written material involving a detailed account of the exact procedures and the sequence in which they are to be used
F2.2 ☐ Formal training session that includes a detailed account of the exact procedures and the sequence in which they are to be used
F2.3 ☐ Written material involving an overview of broad principles and a description of
the intervention phases

F2.4 □ Formal or informal training session involving an overview of broad principles
and a description of the intervention phases

F3. Adaptation procedures are specified (select one) □ yes □ no □ unknown

Rating for Implementation Fidelity for Participant 3 (select 0, 1, 2, or 3): □ 3 □ 2 □ 1 □ 0

Participant 4:

F1. Evidence of Acceptable Adherence (answer F1.1 through F1.3)

F1.1 □ Ongoing supervision/consultation
F1.2 □ Coding intervention sessions/lessons or procedures
F1.3 □ Audio/video tape implementation (select F1.3.1 or F1.3.2):

F1.3.1 □ Entire intervention
F1.3.2 □ Part of intervention

F2. Manualization (select all that apply)

F2.1 □ Written material involving a detailed account of the exact procedures and the
sequence in which they are to be used

F2.2 □ Formal training session that includes a detailed account of the exact
procedures and the sequence in which they are to be used

F2.3 □ Written material involving an overview of broad principles and a description of
the intervention phases

F2.4 □ Formal or informal training session involving an overview of broad principles
and a description of the intervention phases

F3. Adaptation procedures are specified (select one) □ yes □ no □ unknown

Rating for Implementation Fidelity for Participant 4 (select 0, 1, 2, or 3): □ 3 □ 2 □ 1 □ 0

Average Fidelity Rating Across Participants:

\[ \frac{\sum X}{N} \]

X = individual fidelity ratings for each participant
N = number of participants in the study

Overall Rating for Fidelity: (select 0, 1, 2, or 3): □ 3 □ 2 □ 1 □ 0

(Round up or down to the nearest whole number when providing a mean rating for the study. For example, 2.0 to
2.4 rated as 2; 2.5 to 2.9 rated as 3).

G. Replication

G1. Same Intervention (select one of the following)
   - G1.1 Yes
   - G1.2 No
   - G1.3 Unknown

G2. Same Target Problem (select one of the following)
   - G2.1 Yes
   - G2.2 No
   - G2.3 Unknown

G3. Independent Evaluation (select one of the following)
   - G3.1 Yes
   - G3.2 No
   - G3.3 Unknown

Rating for Replication (select 0, 1, 2, or 3): [ ] 3 [ ] 2 [ ] 1 [ ] 0

H. Site of Implementation

H1. School (if it is a school site, select one of the following options)
   - H1.1 Public
   - H1.2 Private
   - H1.3 Charter
   - H1.4 University Affiliated
   - H1.5 Alternative
   - H1.6 Not specified

H2. Non School (if it is a non school site, select one of the following options)
   - H2.1 Home
   - H2.2 University Clinic
   - H2.3 Summer Program
   - H2.4 Outpatient Hospital
   - H2.5 Partial inpatient/day Intervention Program
   - H2.6 Inpatient Hospital
   - H2.7 Private Practice
   - H2.8 Mental Health Center
   - H2.9 Residential Treatment Facility
   - H2.10 Other (specify): __________________________
   - H2.12 Unknown/insufficient information provided
I. Follow Up Assessment

☐ Timing of follow up assessment: specify____________________

☐ Number of participants included in the follow up assessment: specify____________________

☐ Consistency of assessment method used: specify____________________

Rating for Follow Up Assessment (select 0, 1, 2, or 3):  □ 3  □ 2  □ 1  □ 0

III. Other Descriptive or Supplemental Criteria to Consider

A. External Validity Indicators

A1. Sampling procedures described in detail □ yes  □ no

Specify rationale for selection: ___________________________________________

Specify rationale for sample size: __________________________________________

A1.1 Inclusion/exclusion criteria specified  □ yes  □ no

A1.2 Inclusion/exclusion criteria similar to school practice  □ yes  □ no

A1.3 Specified criteria related to concern  □ yes  □ no

A2. Participant Characteristics Specified for Treatment and Control Group

<table>
<thead>
<tr>
<th>Participants from Treatment Group</th>
<th>Grade/age</th>
<th>Gender</th>
<th>Ethnicity or Multi-ethnic</th>
<th>Ethnic Identity</th>
<th>Race(s)</th>
<th>Acculturation</th>
<th>Primary Language</th>
<th>SES</th>
<th>Family Structure</th>
<th>Locale</th>
<th>Disability</th>
<th>Functional Descriptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ Child/Student</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ Parent/caregiver</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ School</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ Child/Student</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ Parent/caregiver</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ School</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ Child/Student</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ Parent/caregiver</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ School</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ Child/Student</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ Parent/caregiver</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ School</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ Child/Student</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ Parent/caregiver</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ School</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☐ Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Procedural and Coding Manual

#### Participants from Treatment Group

<table>
<thead>
<tr>
<th>Grade/age</th>
<th>Gender</th>
<th>Ethnicity or Multi-ethnic</th>
<th>Ethnic Identity</th>
<th>Race(s)</th>
<th>Accultur ation</th>
<th>Primary Language</th>
<th>SES</th>
<th>Family Structure</th>
<th>Locale</th>
<th>Disability</th>
<th>Functional Descriptor</th>
</tr>
</thead>
</table>
| Child/Student
  - Parent/caregiver
  - Teacher
  - School
  - Other
|       |       |                           |                |          |               |                  |     |                 |        |             |                      |
| Child/Student
  - Parent/caregiver
  - Teacher
  - School
  - Other
|       |       |                           |                |          |               |                  |     |                 |        |             |                      |
| Child/Student
  - Parent/caregiver
  - Teacher
  - School
  - Other
|       |       |                           |                |          |               |                  |     |                 |        |             |                      |
| Child/Student
  - Parent/caregiver
  - Teacher
  - School
  - Other
|       |       |                           |                |          |               |                  |     |                 |        |             |                      |

A3. Details are provided regarding variables that:

A3.1 Have differential relevance for intended outcomes  □ yes  □ no

Specify: ____________________________

A3.2 Have relevance to inclusion criteria  □ yes  □ no

Specify: ____________________________

A4. Receptivity/acceptance by target participant population (treatment group)

<table>
<thead>
<tr>
<th>Participants from Treatment Group</th>
<th>Results (What person reported to have gained from participation in program)</th>
<th>General Rating</th>
</tr>
</thead>
</table>
| Child/Student
  - Parent/caregiver
  - Teacher
  - School
  - Other | | □ Participants reported benefiting overall from the intervention |
| Child/Student
  - Parent/caregiver
  - Teacher | | □ Participants reported not benefiting overall from the intervention |

□ Participants reported benefiting overall from the intervention
A5. Generalization of Effects:

A5.1 Generalization over time

A5.1.1 Evidence is provided regarding the sustainability of outcomes after intervention is terminated  □ yes  □ no

Specify: ____________________________________

A5.1.2 Procedures for maintaining outcomes are specified  □ yes  □ no

Specify: ____________________________________

A5.2 Generalization across settings

A5.2.1 Evidence is provided regarding the extent to which outcomes are manifested in contexts that are different from the intervention context  □ yes  □ no

Specify: ____________________________________

A5.2.2 Documentation of efforts to ensure application of intervention to other settings  □ yes  □ no

Specify: ____________________________________

A5.2.3 Impact on implementers or context is sustained  □ yes  □ no

Specify: ____________________________________

A5.3 Generalization across persons

Evidence is provided regarding the degree to which outcomes are manifested with participants who are different than the original group of participants for with the intervention was evaluated  □ yes  □ no

Specify: ____________________________________
B. Length of Intervention (select B1 or B2)

B1. ☐ Unknown/insufficient information provided

B2. ☐ Information provided (if information is provided, specify one of the following:)

B2.1 weeks ______

B2.2 months ______

B2.3 years ______

B2.4 other ______

C. Intensity/dosage of Intervention (select C1 or C2)

C1. ☐ Unknown/insufficient information provided

C2. ☐ Information provided (if information is provided, specify both of the following:)

C2.1 length of intervention session ______

C2.2 frequency of intervention session ______

D. Dosage Response (select D1 or D2)

D1. ☐ Unknown/insufficient information provided

D2. ☐ Information provided (if information is provided, answer D2.1)

D2.1 Describe positive outcomes associated with higher dosage: __________________

E. Program Implementer (select all that apply)

E1. ☐ Research Staff
E2. ☐ School Specialty Staff
E3. ☐ Teachers
E4. ☐ Educational Assistants
E5. ☐ Parents
E6. ☐ College Students
E7. ☐ Peers
E8. ☐ Other
E9. ☐ Unknown/insufficient information provided
F. Characteristics of the Intervener

F1. □ Highly similar to target participants on key variables (e.g., race, gender, SES)
F2. □ Somewhat similar to target participants on key variables
F3. □ Different from target participants on key variables

G. Intervention Style or Orientation (select all that apply)

G1. □ Behavioral
G2. □ Cognitive-behavioral
G3. □ Experiential
G4. □ Humanistic/interpersonal
G5. □ Psychodynamic/insight oriented
G6. □ Other (specify): ___________________
G7. □ Unknown/insufficient information provided

H. Cost Analysis Data (select H1 or H2)

H1. □ Unknown/insufficient information provided
H2. □ Information provided (if information is provided, answer H2.1)

H2.1 Estimated Cost of Implementation: _________________________

I. Training and Support Resources (select all that apply)

I1. □ Simple orientation given to change agents
I2. □ Training workshops conducted

# of Workshops provided ______

Average length of training ______

Who conducted training (select all that apply)

I2.1 □ Project Director
I2.2 □ Graduate/project assistants
I2.3 □ Other (please specify):
I2.3 □ Unknown

I3. □ Ongoing technical support
I4. □ Program materials obtained
I5. □ Special Facilities
I6. □ Other (specify):

J. Feasibility

J1. Level of difficulty in training intervention agents (select one of the following)

J1.1 □ High
J1.2 □ Moderate
J1.3 ☐ Low
J1.4 ☐ Unknown

J2. Cost to train intervention agents (specify if known): ______________________

J3. Rating of cost to train intervention agents (select one of the following)

J3.1 ☐ High
J3.2 ☐ Moderate
J3.3 ☐ Low
J3.4 ☐ Unknown
# Summary of Evidence for Single-Participant Design Studies

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Overall Evidence Rating</th>
<th>Description of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NNR = No numerical rating</td>
<td>Strong Promising Weak No/limited evidence</td>
</tr>
<tr>
<td></td>
<td>or 0 - 3</td>
<td>or Descriptive ratings</td>
</tr>
</tbody>
</table>

## General Characteristics

- General Design Characteristics
- Other Design Characteristics (when randomization is used)
- Type of Program
- Stage of Program
- Concurrent or Historical Intervention Exposure

## Key Features

- Measurement: Reliability and Validity
- Quality of Baseline
- Measures Support Primary/Secondary Outcomes
- Educational/Clinical Significance
- Identifiable Components
- Implementation Fidelity
- Replication
- Site of Implementation
- Follow Up Assessment Conducted
<table>
<thead>
<tr>
<th>Descriptive or Supplemental Criteria</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>External validity indicators</td>
<td></td>
</tr>
<tr>
<td>Length of intervention</td>
<td></td>
</tr>
<tr>
<td>Intensity/dosage</td>
<td></td>
</tr>
<tr>
<td>Dosage Response</td>
<td></td>
</tr>
<tr>
<td>Program Implementer</td>
<td></td>
</tr>
<tr>
<td>Characteristics of the Intervener</td>
<td></td>
</tr>
<tr>
<td>Intervention Style/Orientation</td>
<td></td>
</tr>
<tr>
<td>Cost Analysis Data Provided</td>
<td></td>
</tr>
<tr>
<td>Training and Support Resources</td>
<td></td>
</tr>
<tr>
<td>Feasibility</td>
<td></td>
</tr>
</tbody>
</table>
References


APPENDIX A

Task Force Members and Domains
Members of the Task Force on Evidence-Based Interventions in School Psychology *

Task Force, Chair
Thomas R. Kratochwill

<table>
<thead>
<tr>
<th>Conceptual and Methodological Issues</th>
<th>Liaison Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lynn Fuchs</td>
<td>Dick Abidin (Division 53)</td>
</tr>
<tr>
<td>Jan Hughes</td>
<td>George Batch (NASP)</td>
</tr>
<tr>
<td>Tim Keith</td>
<td>Kristin Hawley (Division 12)</td>
</tr>
<tr>
<td>Joel Levin</td>
<td>Steve Quintana (Cultural Diversity, Division 17)</td>
</tr>
<tr>
<td>Arthur Reynolds</td>
<td>Bruce Wampold (Division 17)</td>
</tr>
<tr>
<td>Sylvia Rosenfield</td>
<td>John Weisz (Division 12)</td>
</tr>
</tbody>
</table>

Multicultural Issues

<table>
<thead>
<tr>
<th>Members</th>
<th>Journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colette L. Ingraham</td>
<td><em>Journal of Educational &amp; Psychological Consultation</em> Emilia C. Lopez, Editor</td>
</tr>
<tr>
<td>Evelyn Oka</td>
<td></td>
</tr>
</tbody>
</table>

School-Based Prevention Programs

<table>
<thead>
<tr>
<th>Members</th>
<th>Journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>George Bear</td>
<td><em>Journal of School Psychology</em> Robert C. Pianta, Editor</td>
</tr>
<tr>
<td>Joe Durlak</td>
<td><em>Psychology in the Schools</em> LeAdelle Phelps, Editor</td>
</tr>
</tbody>
</table>

School-Based Intervention Programs for Social Behavior Problems

<table>
<thead>
<tr>
<th>Members</th>
<th>Journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shane Jimerson</td>
<td><em>School Psychology Quarterly</em> Rik Carl D’Amato, Editor</td>
</tr>
<tr>
<td>Greg Waas</td>
<td><em>School Psychology Review</em> Susan M. Sheridan, Editor</td>
</tr>
</tbody>
</table>

Academic Intervention Programs

<table>
<thead>
<tr>
<th>Members</th>
<th>Qualitative Methods Committee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virginia Berninger</td>
<td>Bonnie Nastasi</td>
</tr>
<tr>
<td>Ed Shapiro</td>
<td>Steve Schensul</td>
</tr>
</tbody>
</table>

Family Intervention Programs

<table>
<thead>
<tr>
<th>Members</th>
<th>Comprehensive School Health Care</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cindy Carlson</td>
<td>Rune Simeonsson</td>
</tr>
<tr>
<td></td>
<td>Ronda Talley</td>
</tr>
</tbody>
</table>

School-Wide and Classroom-Based Programs

<table>
<thead>
<tr>
<th>Members</th>
<th>Research-to-Practice Committee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Craig Frisby</td>
<td>Susan Forman</td>
</tr>
<tr>
<td>Charles Maher</td>
<td>Sandra Thompson</td>
</tr>
<tr>
<td>William Strein</td>
<td></td>
</tr>
</tbody>
</table>

* Note: All members are listed in alphabetical order within domains/groups.
APPENDIX A

Members of the Task Force on Evidence-Based Interventions
In School Psychology

List of Task Force Members (M)/Liaisons (L)

Dick Abidin (L)
University of Virginia
405 Emmet Street, Ruffner Hall
Charlottesville, VA 22903
Phone: (804) 982-2359
Fax: (804) 924-1433
Email: rra@virginia.edu

George Batsche (L)
4925 Londonderry Dr.
Tampa, FL 33647-1333
Phone: (813) 971-5881 (H)
Phone: (813) 974-9472 (W)
Email: batsche@tempest.coedu.usf.edu

George Bear (M)
205 Abbotsford Dr.
Middletown, DE 19709
Phone: (302) 378-4524 (H)
Phone: (302) 831-1645 (W)
Fax: (302) 831-4445
Email: gbear@udel.edu

Virginia Berninger (M)
Center on Human Development & Disabilities
University of Washington
322 Miller, Box 353600
Seattle, WA 98195-3600
Phone: (206) 616-6372
Fax: (206) 543-8439
Email: vwb@u.washington.edu

Cindy Carlson (M)
University of Texas at Austin
Dept Educational Psychology
SZB EDB 504
Austin, TX 78712-1296
Phone: (512) 471-4407
Fax: (512) 459-0977
Email: cindy.carlson@mail.utexas.edu

Rik Carl D’Amato (L)
Division of Professional Psychology & Office of the Dean
College of Education
McKee, 125; Campus Box 106
University of Northern Colorado
Greeley, CO 80639
Phone: (773) 508-2969
Fax: (773) 508-8713
Email: jdurlak@luc.edu

Joseph A. Durlak (L)
Director of Clinical Training
Department of Psychology
Loyola University
6525 N. Sheridan Rd.
Chicago, IL 60626
Phone: (732) 932-1644
Fax: (732) 932-1787
Email: sgforman@rci.rutgers.edu

Susan G. Forman (M)
Rutgers University
18 Bishop Place
New Brunswick, NJ 08901
Craig Frisby (M)
University of Missouri
Educational Psychology/Counseling Psychology
4A Hill Hall
Columbia, MO 65211

Phone: (573) 882-7732
Fax: (573) 884-5989
Email: frisbycl@missouri.edu

Lynn Fuchs (M)
Peabody College, Vanderbilt University
Dept. Special education
Box 328 Peabody
Nashville, TN 37203

Phone: (615) 343-4782
Fax: (615) 343-1570
Email: lynn.fuchs@vanderbilt.edu

Kristin Hawley (L)
University of California
Los Angeles, CA 90095

Phone: (310) 206-7620
Fax: (310) 825-6255
Email: hawley@ucla.edu

Jan Hughes (M)
Department Educational Psychology
College of Education
701 Harrington
Texas A&M University
College Station, TX 77843-4225

Phone: (409) 845-2324
Fax: (409) 862-1256
Email: jhughes@tamu.edu

Colette L. Ingraham (M)
Dept. Counseling & School Psychology
MC 1179, College of Education
San Diego State University
5500 Campanile Drive
San Diego, CA 92182-1179

Phone: (619) 594-6605
Fax: (619) 594-7025
Email: ingraham@mail.sdsu.edu

Shane R. Jimerson (M)
University of California
Graduate School of Education
2208 Phelps Hall
Santa Barbara, CA 93106-9490

Phone: (805) 893-3366
Fax: (805) 893-7264
Email: jimerson@education.ucsb.edu

Tim Keith (M)
University of Austin-Texas
Dept. Educational Psychology
EDB 504
Austin, TX 78712

Phone: (512) 471-0287 or 471-4407
Fax: 
Email: tim.keith@mail.utexas.edu

Thomas R. Kratochwill (Chair)
Department Educational Psychology
University of Wisconsin-Madison
1025 W. Johnson St.
Madison, WI 53706

Phone: (608) 262-5912
Fax: (608) 262-0843
Email: tomkat@education.wisc.edu

Joel Levin (M)
Department Educational Psychology
University of Arizona

Phone: (520) 621-4045
Fax: (520) 621-2909
Appendix

1430 E. 2nd Street
Tucson, AZ 85721
Email: jrlevein@U.arizona.edu

Emilia C. Lopez (L)
Queens College, CUNY
Educational & Community Programs
65-30 Kissena Blvd.
Flushing, NY 11367
Phone: (718) 997-5234
Fax: (718) 997-5248
Email: llopez@cedx.com

Charles Maher (M)
Department of Applied Psychology
Graduate School of Applied & Professional Psychology
Rutgers University, PO Box 819
Piscataway, NJ 08855-0819
Phone: (732) 445-2170
Fax: (732) 445-4888
Email: c Maher@CleIndians.com

Bonnie Nastasi (M)
Associate Director of Interventions
Institute for Community Research
2 Hartford Square West, Suite 100
Hartford, CT 06106-5128
Phone: (860) 278-2044 x237
Fax: (860) 278-2141
Email: bonnastasi@yahoo.com

Evelyn Oka (M)
College of Education
Michigan State University
451 Erickson Hall
East Lansing, MI 48824-1034
Phone: (517) 355-6683
Fax: (517) 353-6393
Email: evoka@pilot.msu.edu

LeAdelle Phelps (L)
409 Baldy Hall
Department of Counseling & Educational Psychology
State University of New York at Buffalo
Buffalo NY 14260
Phone: 
Fax: 
Email: Phelps@acsu.buffalo.edu

Robert C. Pianta (L)
University of Virginia
PO Box 400270
Charlottesville VA 22904-4270
Phone: 
Fax: 
Email: 

Stephen Quintana (L)
Department Counseling Psychology
University of Wisconsin-Madison
1000 Bascom Mall
350 Education Bldg.
Madison, WI 53706
Phone: (608) 262-6987
Fax: (608) 265-3347
Email: quintana@education.wisc.edu

Arthur J. Reynolds (M)
Rm 302 School of Social Work
University of Wisconsin-Madison
Madison, WI 53706
Phone: (608) 263-3837
Fax: (608) 263-3836
Email: ajreynol@facstaff.wisc.edu

Sylvia Rosenfield (M)
Counseling & Personnel Services
Phone: (301) 301-405-2861
University of Maryland  
3214C Benjamin Bldg.  
College Park, MD 20742-1125  
MAIL TO:  
502 Moorings Circle  
Arnold MD 21012

Stephen Schensul, PhD (M)  
Center for International Community Health Studies  
University of Connecticut Health Center  
Farmington, CT 06030-6330

Phone: (860) 679-1570  
Fax: (860) 679-1581  
E-mail: schensul@nso2.uchc.edu

Ed Shapiro (M)  
Lehigh University  
A-215 Mountaintop Campus  
111 Research Drive  
Bethlehem, PA 18015

Phone: (610) 758-3256  
Fax: (610) 758-6223  
Email: Ess2@Lehigh.edu

Susan M. Sheridan (L)  
Department of Educational Psychology  
University of Nebraska-Lincoln  
239 Teachers College Hall  
Lincoln, NE 68588-0345

Phone: (402) 472-6941  
Fax:  
Email: ssheridan2@unl.edu

Elisa Steele Shernoff (M)  
School Psychology Program  
Dept. Educational Psychology  
University of Wisconsin-Madison  
1025 W. Johnson St.  
Madison WI 53706-1796

Phone: (608) 262-0196  
Fax: (608) 265-4559  
Email: ewsteelshern@wisc.edu

Rune Simeonsson (M)  
Professor of Education  
University of North Carolina  
UNC-CH, CB #8185  
Chapel Hill, NC 27599-8185

Phone: (919) 966-6634  
Fax: (919) 966-0862  
Email: rune_simeonsson@unc.edu

William Strein (M)  
Associate Professor and Director  
School Psychology Program  
University of Maryland  
3212 Benjamin Bldg.  
College Park, MD 20742-1125

Phone: (301) 405-2869  
Fax: (301) 405-9995  
Email: strein@umd.edu or ws30@umail.umd.edu

Ronda Talley (M)  
Executive Director and Professor  
Rosalynn Carter Institute for Human Development  
Georgia Southwestern State University  
800 Wheatley Street  
Americus, GA 31709

Phone: (229) 928-1234  
Fax: (229) 931-2663  
Email: TalleyRC@rci.gsw.edu
Sandra Thompson, PhD (M)
4396 N. Atfield Street
Tucson AZ 85719

Phone: (520) 743-3899
Cell: (520) 405-1402
Email: sandithompson@comcast.net

Greg Waas (M)
Department of Psychology
Northern Illinios University
DeKalb, IL 60115-2892

Phone: (815) 753-3508
Fax: (815) 753-8088
Email: gwaas@niu.edu

Bruce Wampold (L)
Department Counseling Psychology
University of Wisconsin-Madison
321 Education Bldg.
Madison, WI 53706

Phone: (608) 262-2878
Fax: (608) 265-3347
Email: wampold@education.wisc.edu

John R. Weisz (L)
University of California
Los Angeles, CA 90032

Phone: (310) 206-7620
Fax: (310) 825-6255
Email: weisz@psych.ucla.edu

M = Member
L = Liaison

(TaskForce.Members.List.doc)
Appendix B

General Guidelines for Conducting a Literature Review

Prepared by

Joseph A. Durlak, PhD
General Guidelines for Conducting a Literature Review

1. **Formulate a good research question**: A “good” question is specific, potentially answerable, and important (conceptually, theoretically, or practically).

2. **Define the relevant research domain**: Develop specific inclusionary and exclusionary criteria.

   Key here: Does the research domain have good construct validity? Are you delineating an area that most would find acceptable, or is it too broad, too narrow, or exclusive of some important parameter?

   Are your inclusionary and exclusionary criteria defensible, workable, and clear? These criteria should make it clear which types of studies you are and are not examining. Typically, these criteria specify the relevancy of types of designs, populations, interventions, outcomes, time frame, locale, source of publication, and language.

3. **Obtain a representative set of studies for review**: Impossible to do an absolutely comprehensive review in the sense of locating every study ever done. Idea is to avoid a biased search in which you only obtain the easy-to-find studies. Try to obtain a representative set of investigations.

   Use multiple strategies to locate studies:
   
   - Manual search of relevant journals
   - Computer searches using multiple data bases
   - Inspect reference lists of previous research reviews and each identified report
   - Contact authors who have recently published in the area
   - Contact organizations who fund or publish relevant work
   - Do not restrict search to published studies. Please note that the Task Force, like other work groups identifying effective interventions, will restrict itself to published work although a complete literature review typically includes unpublished reports.
Example Literature Search

Purpose of the Review

The purpose of this review is to assess the amount and quality of the evidence that exists on the outcomes of different types of preventive interventions addressing internalizing problems. Specifically, we will apply the Division 16 Task Force criteria to evaluate different interventions. We hope to indicate: (a) which interventions seem to have the strongest empirical support on different criteria; and (b) which interventions have been investigated and lack sufficient data so that it is not possible to reach any conclusions one way or another regarding their impact on different criteria. (This is just another way of saying: which programs work, which do not work, and for which are there insufficient data to reach any conclusions one way or another).

Reviewers

The main reviewers will consist of Joseph A. Durlak from Loyola University Chicago and Patricia Gryczyk, from the Institute of Juvenile Research in Chicago. These individuals may also be assisted by one or two graduate students.

Domain Definition

We will include both universal (primary) and indicated (secondary) prevention programs. Universal prevention programs are those targeted at entire populations, e.g., all junior high students, all first and second graders, all high school students, and so on. Indicated prevention programs are defined as those that conduct some systematic population-wide screening processing to identify mild to moderately serious internalizing conditions and then offer prompt intervention to those who met the screening criteria. Children with serious (i.e., clinical-level) or chronic problems meriting treatment will be excluded. The process to identify those with incipient problems must be systematic. Usually, this is done by screening a specific school population either through interviews or more commonly with self-report measures to identify eligible candidates. Only those meeting the screening criteria are offered intervention.

Occasionally, multi-gating procedures are used to ascertain the nature and level of problems (e.g., Clarke et al., 1995; Stark et al., 1987). Programs that merely ask teachers to refer students whom they think qualify will be excluded unless there is some separate assessment to confirm the presence of the target problem. That is, teachers might be asked to nominate up to three “shy” children, or children who are displaying behavioral problems in class. Unless the researchers separately confirm the existence of problems (e.g., through sociometrics or behavioral observations), however, these reports will not be included. Admittedly, there is likely to be some overlap between what we may call indicated prevention and what others might consider treatment for established problems. Hopefully, any overlap that does exist can be identified as we each proceed with different reviews.

Inclusionary and Exclusionary Criteria

Types of designs. We will include between-group studies that contain some type of control or comparison group and single participant designs. No one group pre-post designs will be included.

Populations. All school-age groups will be included (i.e., preschool through high school). Aside from a determination if the presenting problem represents a clinical-level disorder or a chronic condition, those with both mild and moderate difficulties will be included.

Types of problems included. We will focus on the following problems: depression, anxiety, test and performance anxiety, shyness and social isolation, and somatic problems. Somatic problems will include eating disorders. Children with more than one type of problem will be included as long as one of the above problems is present. We will also include suicide prevention programs because these typically target depression and suicidal thoughts as outcomes.
Years of search and publication status. Published reports appearing from 1980 to December 2002 will be included. Unpublished studies will be excluded.

Outcomes. Only interventions that include some assessment of the target symptom/problem will be included. Programs that only assess knowledge or attitudes will be excluded.

Language and setting. The report must appear in English, but no restrictions as to country are imposed. Therefore, interventions conducted throughout the world are eligible.

Search Procedures

1. The following search procedures will be used. Computer-based searches of PSYC INFO and MEDLINE will be conducted using a variety of search terms. Iterative search procedures will be used. That is, some search terms will be employed, entries will be inspected, and search terms will be trimmed and expanded to obtain more “hits.” Reference librarians will be consulted to assist in identifying the idiosyncratic features of different computer data bases.

2. The reference lists of published reviews will be consulted to identify studies (e.g., Compas, Conner, & Wadsworth, 1997; Donovan & Spence, 2000; Durlak & Wells, 1997, 1998; Gilham, Shatte, & Freres, 2000; Greenberg, Domitrovich, & Bumbarger, 2001; Mazza, 1997; Miller & DuPaul, 1996).

3. The reference lists of each individual report as it is obtained will also be examined.


5. We will contact authors who have published relevant prevention studies in the past two years (2001 and 2002) to ask them for copies of any in-press outcome studies they have completed.

Based on our preliminary knowledge of this research area, we expect to find approximately 50-70 relevant investigations for review.
Example of Computer Database Search Procedures for the Prevention of Internalizing Disorders in Children and Adolescents

SEARCH USING PSYCINFO DATABASE

Here is an example of how the PsycInfo database could be searched for relevant studies. The same general procedures usually apply to other databases such as MEDLINE, although the specific keywords used by the database might differ. A good resource for additional information on searching computer databases is Reed and Baxter (1994).

Step 1: Enter keyword or phrase

- Start with the keyword “internalizing”
- The database maps this term to a subject heading, which you have to choose
- The closest term is “internalization”, so click on this and continue
- This search yields 579 results! Now we want to narrow this down

Step 2: Narrow down the search

- Specify limits for relevant variables (e.g., language, age groups)
  - By clicking on “limits” (at the top of the screen), you can then choose a search you’ve done and limit the search according to certain age groups, language, etc.
  - Select the search for “internalization”, and scroll down to specific limit fields (age group, publication types, population groups, etc.). Then highlight relevant age groups; in this case, all childhood and adolescent codes are relevant. Also, highlight English for language, if desired.
  - After selecting these limits and clicking the “limit search” button, this search yields 304 results. This is still a lot of references, but we can now begin to pick out the more relevant ones.

Step 3: Finding more specific search terms

- What keywords and search terms are relevant studies classified by?
  - After perusing through the 304 results, we come across several relevant articles and reviews. By clicking on the “complete reference” link for these articles, we are able to find out what search terms PsycInfo has applied to these specific articles.
    - Examples: We find a review on prevention of childhood anxiety disorders by Donovan and Spence (2000). After clicking on the “complete reference” link, we find that the keywords for this review are: anxiety disorders; primary mental health prevention. After doing the same thing for the review on preventing depression by Compas et al. (1997), we find that the keywords are: major depression; primary mental health prevention.
    - Reading other full references of relevant articles yields the following additional keywords: attempted suicide, suicide prevention, emotionally disturbed, anxiety, anxiety disorders, anxiety management, depression, self-esteem, stress management, panic disorder, adjustment, and prevention.

Step 4: Use more specific key words to perform searches
Now that we know what keywords will yield effective searches, we can use these to find relevant references:

- To find articles about preventing depression in children and adolescents, we first start by typing in “depression” as the keyword.
- This keyword is then mapped to different subject headings, and we choose both “depression (emotion)” and “major depression” since these are both relevant to the search. This search yields 52,342 matches, so we need to narrow it down further.
- We apply limits for the English language and childhood and adolescent age groups as we did previously, and now the results number 6,998.
- However, we only want to look at studies about preventing depression, so we need to first look up the references for prevention and then combine the two searches. After entering “prevention” as a keyword, this term is then mapped to “prevention” and “primary mental health prevention” since both terms came up as relevant keywords.
- We then limit this search in the same way, by choosing the English language and childhood and adolescent age groups, and the results for this search number 5,247.
- We then combine the search for depression and the search for prevention by clicking on the “combine” icon at the top of the screen and selecting these two searches to be combined. The results of this combined search yield 119 references dealing with prevention of depression in children and adolescents.
- This same technique is repeated with “anxiety” (mapped to “anxiety”, “anxiety management”, “anxiety disorders”) and combined with prevention (“prevention” and “primary mental health prevention”), and this search yields 82 results.
- We then repeat this technique with the other keywords that were identified above.

**Step 5: Other helpful hints**

- Searching by author: When you find a relevant article, search by the name of the first, second, and even third author to find all articles written by the authors (you can this by clicking on the “author” icon at the top of the screen). Since researchers usually do research in a limited number of areas, this process will usually yield additional relevant studies that the previous searches may not have generated.
- Print out the search history (see attached sheets). This tactics helps you to remember the terms you have used in your search, which may be helpful at a later time if you come across new keywords to use.
APPENDIX C

General Resources
General Resources on Evidence-Based Interventions


General Resources on Conducting Literature Reviews


General Resources on Group and Single-Participant Design Meta-Analysis


APPENDIX D

Supplemental Coding Options for Quasi-Experimental Group-Based Designs
QUASI-EXPERIMENTAL DESIGNS WITHOUT CONTROL GROUPS

- **The One-Group Posttest-Only Design**
  \[ X \quad O_1 \]

- **The One-Group Posttest-Only Design With Multiple Substantive Posttests**
  \[ X_1 \quad \{O_{1A}, O_{1B}, \ldots, O_{1n}\} \]

- **The One-Group Pretest-Posttest Design**
  \[ O_1 \quad X \quad O_2 \]

- **The One-Group Pretest-Posttest Design Using a Double Pretest**
  \[ O_1 \quad O_2 \quad X \quad O_3 \]

- **The One-Group Pretest-Posttest Design Using a Nonequivalent Dependent Variable**
  \[ \{O_{1A}, O_{1B}\} \quad X \quad \{O_{2A}, O_{2B}\} \]

- **The Removed-Treatment Design**
  \[ O_1 \quad X \quad O_2 \quad O_3 \quad \times \quad O_4 \]

- **The Repeated-Treatment Design**
  \[ O_1 \quad X \quad O_2 \quad O_3 \quad \times \quad O_4 \]
QUASI-EXPERIMENTAL DESIGNS THAT USE COMPARISON GROUPS AND PRETESTS

- **Untreated Control Group Design with Dependent Pretest and Posttest Samples**

  \[
  \begin{align*}
  \text{NR} & \quad O_1 \times O_2 \\
  \text{NR} & \quad O_1 \times O_2 
  \end{align*}
  \]

- **Untreated Control Group Design with Dependent Pretest and Posttest Samples Using a Double Pretest**

  \[
  \begin{align*}
  \text{NR} & \quad O_1 \quad O_2 \times O_3 \\
  \text{NR} & \quad O_1 \quad O_2 \times O_3 
  \end{align*}
  \]

- **Untreated Control Group Design with Dependent Pretest and Posttest Samples Using Switching Replications**

  \[
  \begin{align*}
  \text{NR} & \quad O_1 \times O_2 \quad O_3 \\
  \text{NR} & \quad O_1 \quad O_2 \times O_3 
  \end{align*}
  \]

- **Untreated Control Group Design with Dependent Pretest and Posttest Samples Using Reversed-Treatment Control Group**

  \[
  \begin{align*}
  \text{NR} & \quad O_1 \times O_2 \\
  \text{NR} & \quad O_1 \times O_2 
  \end{align*}
  \]

- **Cohort Control Group Design**

  \[
  \begin{align*}
  \text{NR} & \quad O_1 \\
  \text{NR} & \quad X \quad O_2 
  \end{align*}
  \]

- **Cohort Control Group Design with Pretest from Each Cohort**

  \[
  \begin{align*}
  \text{NR} & \quad O_1 \quad O_2 \\
  \text{NR} & \quad O_3 \quad X \quad O_4 
  \end{align*}
  \]