CODING MANUAL FOR

SINGLE-PARTICIPANT INTERVENTION RESEARCH DESIGN

Task Force on Evidence-Based Interventions in

School Psychology
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 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”).
A2. General Study Characteristics and Cultural Factors.

As with group-based designs, investigators address the foundations for the study and the logical connections among the theoretical-empirical basis, research questions, and methodology (here, we limit discussion to single-participant designs). Note that these designs may be amenable to addressing the cultural factors given investigators often have considerable contextual information due to intense study of units (e.g., students) within operationally defined environments (e.g., classrooms). The degree to which such factors can be addressed may well be a function of what research and consultation models (e.g., conjoint-behavioral or participatory) were used to envisage the questions to be addressed by the design.

With regard to cultural diversity considerations, this section addresses the rationale and appropriateness of the study’s conceptual basis, questions, and methods in terms of the target population. Issues to consider include:

- The extent to which the theoretical basis and key constructs for the study were grounded in etic (researcher) and emic (participant) perspectives; for example, the study is based on an integration of both existing theory and formative research to test the relevance of existing theory to the cultural ecology and experiences of the target population. In a single-participant study, this consideration may be manifested as a clear effort to address emic perspectives about a given behavior or outcomes.

- The rationale for the theoretical base and key constructs was explicit; that is, the link of researcher’s and participants/stakeholders’ perspectives to the conceptual basis of the study were clearly stated. This may appear as an explicit effort to address contrasting values among stakeholders (e.g., a rationale, teacher, student and interventionist may have different values around a given outcome, and the researcher both recognizes these differences and addresses them when describing impacts).

- The research questions reflect the views of the researcher and participants (etic-emic views) and guide the selection of methodology and sample. For example, an investigator may choose a single-participant design on the basis that it is amendable to developing operational definitions of an outcome embedded in context (or at least highlight the benefits of this approach).

Wolf (1978) brought the concept of social validity into behavioral technology. He stated that traditional means of evaluating treatment solely in terms of measured outcomes was no longer sufficient. He proposed that we need to be concerned with socially relevant measures in addition to objective measures. Wolf suggested that society has to validate our work on at least three levels: (1) the social significance of the goals, (2) the social appropriateness of the procedures, and (3) the social importance of the effects. Collectively he referred to these judgments as social validity.

A2.1 Social Significance of Goals – This criterion refers to the assessment of the importance of specific variables in behavior change. Kazdin (1977) proposed two procedures: social comparison and social evaluation. Social comparison involves the documentation of the skills displayed by normal children. These are then considered socially appropriate educational goals. Social evaluation involves both formal and informal methods of asking the opinion of others concerning the social appropriateness of treatment goals. Typically it is the opinion of experts.

A2.2 Social Appropriateness of Procedures – This criterion is concerned with ethics, cost, and practicality. Research in this area has focused on the acceptability of behavioral interventions as operationalized by Kazdin. If a treatment is not found to be acceptable it is may be less likely to be adopted and as a result won’t be effective.

A2.3 Social Importance of Effects – This criterion refers to the judgment of whether the consumer is satisfied with the results yielded by the intervention. Wolf contends that because behavioral
treatments are designed to help someone with a problem, whether or not the program is helpful can only be assessed by the consumer. Objective data are not important, if the consumer is satisfied then treatment is said to produce socially important changes.

A3. Theoretical basis. Intervention design is based upon a conceptual model that is grounded in theory and applied to the empirical study of the target phenomenon and participants. The rating on the item reflects the extent to which the theoretical basis is described sufficiently and guides (is linked to) the intervention.

3 = Described theoretical basis and clearly linked to the intervention.
2 = Described theoretical basis, but no connection to the intervention.
1 = Provided some reference to theory but lacks clear conceptual model and its relation to the intervention.
0 = Theoretical basis absent.

A4. Documents the relationship between the implementers and participants. Description provided regarding procedures and interpersonal processes to establish the relationship for delivering the intervention; for example, nature of the relationship, processes or procedures used to continue pre-existing relationship or establish new relationship, develop and maintain rapport and trust, monitor interpersonal relationships, and facilitate effective working relationships.

3 = Detailed description regarding the interpersonal processes used to establish and maintain the relationship between implementer and participants.
2 = Detailed description of relationship development procedures, but lacking detail on some aspects of the relationship processes.
1 = Provides overview of relationship development procedures and processes, but lack details.
0 = No description of relationship processes provided.

A5. Ecological validity. This characteristic refers to the extent to which the relevance or appropriateness of the study’s empirical-theoretical (conceptual) basis has been established for the study participants. To receive the maximum rating, ecological validity must be established through formative research of the design context or a consultation model that addresses whether outcomes are culturally relevant and socially valid.

3 = Ecological validity of conceptual base established through formative research with study participants
OR conceptual base is culturally derived.
2 = Existing theory adapted based on research with study participants.
1 = Existing theory adapted based on empirical evidence with similar or related populations.
0 = Existing theory used without attention to ecological validity.

A6. Researcher perspective. This criterion refers to the extent to which the researcher has articulated personal beliefs (consider aforementioned etic and emic perspectives) related to the conceptualization or empirical-theoretical basis of the study.

3 = Researcher articulated personal beliefs and relation to the conceptualization of the study.
2 = Researcher articulated personal beliefs with vague link to conceptualization.
1 = Researcher articulated personal beliefs with no link.
0 = No discussion of researcher’s beliefs related to the study’s conceptualization.

A7. Moderator variables. Includes the perceptions of the participants regarding the acceptability and cultural validity of the intervention; that is, whether the intervention content and process is consistent with their cultural perspectives and experiences, and whether they find the intervention to be feasible and enjoyable, and beneficial for their daily lives. These considerations, in conjunction with conditions for
implementation, are critical for and interpreting the outcomes, and determining the likelihood of sustainability and transferability of interventions.

3 = Investigated effects of cultural variables, with culture operationalized with reference to acculturation, perceptions of power, oppression.
2 = Investigated effects of cultural variables, with culture operationalized as demographic variables (e.g., race, ethnicity, language).
1 = Interpreted results in terms of culture without formal investigation, but with reference to other research with similar or related cultural groups.
0 = No evidence that moderator effects were investigated and no reference to influence of culture in interpretation of findings.

A8. Rival interpretations. In presenting the conceptual basis of the study, the researcher evaluates rival cultural, methodological, statistical, or theoretical explanations and hypotheses. This criterion refers to the extent to which alternative explanations and interpretations of the study’s conceptual foundations and findings are considered by the researchers. This process can be facilitated through interaction with professional colleagues who are not involved in the study (peer review) and/or through participation of the stakeholders throughout the research process. The researcher makes this process explicit by discussing the process and outcome for considering rival interpretations, including how the process was conducted, the alternative perspectives that were considered, and how final decisions were reached.

3 = Investigated multiple rival hypotheses (e.g., cultural, methodological, statistical, and/or theoretical explanations) and provided empirical evidence.
2 = Investigated one rival hypothesis (e.g., cultural, methodological, statistical, and/or theoretical explanations) and provided empirical evidence.
1 = Considered rival explanations without empirical evidence.
0 = No evidence that rival hypotheses were investigated or considered (examined data only from one perspective).

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.

Interactive process and culture. These criteria are concurrent with the aforementioned questions regarding extent of engagement and can have impacts on whether culturally-nuanced information is captured and accounted. The level of engagement affects both (a) the depth and reliability of representations, and (b) the scope and depth of understandings and relationships within the study. High levels of engagement assure deep and accurate representations and understandings, thereby affecting understanding and reliability in the study of intervention outcomes. Prolonged engagement means that the data are collected over a substantial period of time to ensure accuracy of representations. Persistent observation means that the data collection is progressively focused to ensure thorough understanding of consistency and variation likely to occur in intervention process and outcomes. The goal is for researchers to collect data in a manner that guarantees sufficient scope and depth through prolonged engagement and persistent observation. For example, a researcher who comes in one time and collects data would get a lower score whereas a researcher who is deeply involved over a long period of time so that behaviors and thoughts of participants are represented would get a high score. The goal is to attain validity and reliability through trustworthiness.

In the context of single-participant designs, prolonged engagement and persistent observations can be thought of in terms of having an adequate number of data points in each phase of the design and continuous assessment. If the investigator utilizes these features in part to establish an understanding of cultural factors (this may be referred to as context) then it may be appropriate to afford related credit.
3 = Provided evidence for high level of engagement to ensure deep and accurate representation.
2 = Provided evidence for some level of engagement to ensure deep and accurate representation.
1 = Provided evidence for minimal level of engagement to ensure deep and accurate representation.
0 = Provided no evidence for level of engagement to ensure deep and accurate representation.

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.

Cultural Moderator Variables. Investigated effects of cultural moderating variables on the outcome. Attention to moderator variables includes consideration of the cultural variables that may impact program implementation and outcomes. Moderator variables include the perceptions of the participants regarding the
acceptability and cultural validity of the intervention; that is, whether the intervention content and process is consistent with their cultural perspectives and experiences, and whether they find the intervention to be feasible and enjoyable, and beneficial for their daily lives. These considerations, in conjunction with conditions for implementation, are critical for and interpreting the outcomes, and determining the likelihood of sustainability and transferability of interventions.

3 = Investigated effects of cultural variables, with culture operationalized with reference to acculturation, perceptions of power, oppression.
2 = Investigated effects of cultural variables, with culture operationalized as demographic variables (e.g., race, ethnicity, language).
1 = Interpreted results in terms of culture without formal investigation, but with reference to other research with similar or related cultural groups.
0 = No evidence that moderator effects were investigated and no reference to influence of culture in interpretation of findings.

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).

Cultural Significance

This criterion refers to the cultural relevance or significance (i.e., cultural validity) of key outcome measures. *Culture* is defined as the shared norms, values, beliefs, practices, and behaviors relevant to the target population; and reflects cultural experiences that extend beyond the traditional categories of race, ethnicity, and language to encompass the shared experiences of any group (e.g., culture of classroom, school, neighborhood, family, peer group). The selection of key outcome variables and indicators is based on formative research that documents relevance to ecological/social contexts and the cultural experiences of the target population and the participants. The indicators are meaningful and ecologically valid given the context and experiences of the participants. The indicator has relevance to the real-life experiences of the target population and the participants as supported by research evidence. For example, in research on prevention of sexually transmitted diseases, formative research has established that condom use is a viable and culturally-appropriate alternative for preventing HIV infection in the target population and for the specific participants. Thus, the indicator represents concepts that have legitimacy within the shared social-cultural experiences of the target population and participants. Ratings for this criterion are based on the extent to which the relevance to the target population has been established and the nature of the evidence.

3 = Investigated effects of cultural variables, with culture operationalized with reference to acculturation, perceptions of power, oppression.
2 = Investigated effects of cultural variables, with culture operationalized as demographic variables (e.g., race, ethnicity, language).
1 = Interpreted results in terms of culture without formal investigation, but with reference to other research with similar or related cultural groups.
0 = No evidence that moderator effects were investigated and no reference to influence of culture in interpretation of findings.

Research Methodology

Research methodology pertains to the larger conceptualization of the research used to establish the effectiveness of an intervention and specific issues regarding the methods used to document and report
intervention effects. This domain includes topics associated with the research design, sampling, selection of measures, descriptions of participants and researchers, and qualities of the measures to establish the effects of an intervention. The conceptualizations of the researchers determine the research questions posed, definitions of constructs, methods of data collection, descriptions of participants, and relationships between research and participants, and measurements used. Studies will be evaluated with regard to the qualities of the conceptualizations and measures used to establish the effects of an intervention. Coders will rate the extent to which there is evidence for each of the characteristics or features within the research methodology section. Coding rubrics are delineated for items A1 to A5, following this general progression:

To be rated a 3 for strong evidence, studies must provide evidence for all elements of the criterion specified in this item. Refer to the Coding Protocol for specific elements of this criterion that must be included.

To receive a rating of 2 for promising evidence, evidence must be provided for most of the elements of the criterion specified in this item.

A rating of 1 for weak evidence would require some elements, but not comprehensive or inclusive of most.

A rating of 0 indicates that no evidence of the criterion is available or that the evidence provided is seriously insufficient in the required characteristics or features.

**A1. Characteristics of the data collector(s).** Studies will be evaluated regarding the level of similarity between data collectors and target participants and the cultural competence of the data collector with respect to the specific target populations.

3 = Provided evidence for high level of similarity to target participants that goes beyond proxy variables (e.g., rapport, trust, communication, ethnic identity) and cultural competence documented.

2 = Provided evidence for some level of similarity to target participants that goes beyond proxy variables (e.g., rapport, trust, communication, ethnic identity) and cultural competence documented.

1 = Provided no evidence for similarity on these variables, but cultural competence is documented.

0 = Provided no evidence for similarity to target participants and/or cultural competence.

**A2. Characteristics of participants.** Reviewers will evaluate the clarity of descriptions of what culture means for specific groups and the measurement of individuals’ psychological experience or interpretation of cultural membership. For example, to what extent does the study define how culture is conceptualized by participants and how participants interpret their cultural membership? For example, are psychological experiences, such as acculturation, perceptions of power, and oppression operationalized and measured among participants?

3 = Clear definition of what culture means for groups and measurement of individuals’ psychological experience/interpretation of cultural memberships (e.g., acculturation, perceptions of power, oppression).

2 = Clear definition of what culture means for groups without measurement of individuals’ psychological experience/interpretation of cultural memberships individuals’ interpretations (Culture is conceptualized but not operationalized).

1 = Equates culture with demographic labels.

0 = No attempt to describe, define and/or measure culture.

**A3. Sample appropriate research methods.** Reviewers will rate the extent that research methods guide sampling procedures (establishing clear links between research methods and sampling), and sampling
approaches are appropriate to the research methods.

3 = Clear links established between research methods and sampling, and sampling is appropriate to the research methods.

2 = Vague or no links established between research methods and sampling, but sampling is appropriate to the research methods.

1 = Links established between research methods and sampling, but sampling is inappropriate to the research methods.

0 = No links are established and sampling is inappropriate to research methods.

A4. Operationalization. Specifying the link between key abstract constructs (variables) and data collection methods (operations). Evaluate the clarity of the relationship between key abstract constructs or variables and the data collection methods or operations and the degree that they are operationalized.

3 = Clear links established between constructs and methods, and all key constructs are clearly operationalized.

2 = Some, but not all, key constructs are clearly operationalized.

1 = Vague reference to link between constructs and methods.

0 = No evidence that key constructs are operationalized.

A5. Integration of data from multiple sources, methods, and investigators. Extent to which studies used multiples of all three: sources, methods, and investigators.

3 = Used multiple sources, methods, and investigators.

2 = Used two of the following: multiple sources, multiple methods, multiple investigators.

1 = Used one of the following: multiple sources, multiple methods, multiple investigators.

0 = No evidence of multiple sources, methods, or investigators.

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.

Cultural Appropriateness of the Measures

Many of the aforementioned concerns regarding cultural validity of measures (see section of Group-Based Designs) apply here. Studies will be evaluated on the extent to which evidence is provided to demonstrate the cultural appropriateness of the measures for the study participants and the methods by which this determination was made. In rating this item, consider the following dimensions: meaning, language, dialect, and response format used in the methods of data collection and responses. For example evaluate evidence that participants provided data in a way that is culturally familiar and in their specific language and dialect.

Record the extent to which there is direct empirical evidence that the measures are in a format that is culturally appropriate for this specific target group versus evidence from related or other groups. A maximum rating for cultural appropriateness indicates that measures were developed through formative research with the study participants. Note that when the outcome measure entails observing behavior bound to a given context, cultural validity may be assumed so long as an effort is made to check with stakeholders to assess that a given behavior (e.g., using public pay phones) takes on an entirely new meaning with some subgroups (e.g., Old Order Amish), see Rogler (1999). See sources such as Castillo (1997), Leong & Lau (2001), Snowden (2001) and the U.S. Department of Health and Human Services (2001) for examples of how given behaviors can take on different meanings as a function of cultural influences.
3 = Developed measure for use with study participants on the basis of empirical evidence (conducted formative research and developed measure).
2 = Adapted existing measure for use with study participants on the basis of formative research and/or empirical evidence with study participants.
1 = Developed or adapted measures for use with study participants based on empirical evidence with similar or related populations.
0 = Measure not tailored specifically for study participants.

Cultural Significance

This criterion refers to the cultural relevance or significance (i.e., cultural validity) of key outcome measures. Again, *Culture* is defined as the shared norms, values, beliefs, practices, and behaviors relevant to the target population; and reflects cultural experiences that extend beyond the traditional categories of race, ethnicity, and language to encompass the shared experiences of any group (e.g., culture of classroom school, neighborhood, family, peer group). The selection of key outcome variables and indicators is based on formative research that documents relevance to ecological/social contexts and the cultural experiences of the target populations and the participants. The indicators are meaningful and ecologically valid given the context and experiences of the participants. The indicator has relevance to the real-life experiences of the target population and the participants as supported by research evidence. In single participant designs, the outcome measure and intervention can be culturally intertwined. Therefore, the aforementioned issues of social validity apply here.

3 = Conducted formative research or provided empirical evidence with the target population to show relevance of key outcome variables and indicators to target population.
2 = Cited empirical evidence for relevance of key outcome variables and indicators to target population.
1 = Cited empirical evidence with similar or related populations for relevance of key outcome variables and indicators to target population.
0 = No empirical evidence cited for relevance of outcome variables and indicators to population.

Measures of key outcomes are linked to the conceptual model. The selection of key outcome variables and indicators are linked to the conceptual model (grounded in theory and research). The connections between outcome indicators and variables in the conceptual model are clearly articulated. Outcome indicators reflect different points in the pathways of change (as articulated in the conceptual model) such as independent, mediating, and dependent variables. For example, in the conceptual model, the independent variable (introduction of some treatment such as a behavior plan) is linked to the mediating variable (perceptions of the treatment) that is linked to the dependent variable (observed behavioral change). Indicators for all three variables should be included in order to test proposed pathways of change.

3 = Clear links established between the conceptual model and key outcome indicators.
2 = Some, but not all, key outcomes are clearly linked to conceptual model.
1 = Vague reference to links between key outcomes and conceptual model.
0 = No evidence that key outcomes are linked to conceptual model.
<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></td>
<td>Analyze the relationship among parts of the test and between the test and other constructs/variables</td>
</tr>
<tr>
<td></td>
<td></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>
</tbody>
</table>
### Reliability (*consistency*)

<table>
<thead>
<tr>
<th>Type</th>
<th>Content</th>
<th>Time Interval</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test-retest</td>
<td>Identical</td>
<td>Varies</td>
<td>Give identical instrument twice</td>
</tr>
<tr>
<td>Equivalent</td>
<td>Different</td>
<td>None</td>
<td>Give two forms of instrument</td>
</tr>
<tr>
<td>Internal consistency</td>
<td>Different</td>
<td>None</td>
<td>Divide instrument into halves and score each or use K-R</td>
</tr>
<tr>
<td>Observer agreement</td>
<td>Identical</td>
<td>None</td>
<td>Compare scores obtained by two (or more) observers</td>
</tr>
</tbody>
</table>


### 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 (or the researcher should have used all analyses methods to account for it).

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, all analyses of all primary and secondary outcome measures. Outcomes will be considered statistically significant if they reach a p value 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** (e.g., A / B / A / B, A / B+C / A / B+C).

**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.** Socially 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 3 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 3. 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
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 of 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 opposition 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.

G3. Transferability of intervention. Level of detail provided to permit decisions about transferability (including decisions about similarity of context, participants, intervener, and available resources; key characteristics of the intervention are transportable). Decisions about transferability (external validity) of the intervention are dependent on “rich description” of the study’s conditions. Rich description involves in-depth detailing of the procedures and conditions surrounding implementation – including delivery of the intervention, training of facilitators, monitoring implementation, and making necessary adaptations. The documentation of the procedures and conditions surrounding implementation and adaptation must be of sufficient detail to permit other researchers and practitioners to transfer the intervention to other contexts and participants, and to determine the applicability of fit to intended contexts and participants. Thick description should include information about how decisions regarding adaptation to local conditions were made during the course of the intervention study. Decisions about transportability or “generalization” are ultimately within the purview of the consumer; however, the researcher’s responsibility is to provide sufficient detail to facilitate the consumer’s decision making about applicability and provide a process for adapting the intervention to local conditions.

3 = Complete and detailed description of the context within which the intervention occurs.
2 = Detailed description of some but not all contextual components.
1 = Provides overview of contextual components but lack details.
0 = No description of context.

G4. Participant Perception of Benefits of the Intervention (Receptivity/acceptance by target participant population). Investigated perceptions of benefit for all participant groups and reported participants benefiting overall from the intervention. 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 from the intervention.

3 = Provided evidence of perceived benefits from the intervention for all participant groups.
2 = Provided evidence of perceived benefits from the intervention for some participant groups.
1 = Provided evidence that participants did not perceive benefits from the intervention.
0 = Did not investigate participants perceptions of benefits.

G5. 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 was 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 that 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 (e.g., a video parent training program that requires videotape vignettes be shown to a group of parents is followed by a psychologist facilitated discussion within a school context with 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.
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 Sampling procedures described.**
**A1.2 Rationale for sample selection specified.**
**A1.3 Rationale for sample size specified.**
**A1.4 Evidence provided that sample represents the target population.**
**A1.5 Recruitment procedures are congruent with target cultural group.** Reviewers will determine whether culturally appropriate ways were used to contact, recruit, inform, and maintain participation of the participants.

**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. 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):

A2. General Study Characteristics and Cultural Factors

A2.1. Social Significance of Goals
A2.2 [ ] Social Appropriateness of Procedures
A2.3 [ ] Social Importance of Effects

A3. Rating for Theoretical Basis (select 0, 1, 2, or 3): [ ] 3 [ ] 2 [ ] 1 [ ] 0

A4. Rating for Documenting the Relationship Between the Implementers and Participants (select 0, 1, 2, or 3): [ ] 3 [ ] 2 [ ] 1 [ ] 0

A5. Rating for Ecological Validity (select 0, 1, 2, or 3): [ ] 3 [ ] 2 [ ] 1 [ ] 0

A6. Rating for Researcher Perspective (select 0, 1, 2, or 3): [ ] 3 [ ] 2 [ ] 1 [ ] 0

A7. Rating for Moderator Variables (select 0, 1, 2, or 3): [ ] 3 [ ] 2 [ ] 1 [ ] 0

A8. Rating for Rival Interpretations (select 0, 1, 2, or 3): [ ] 3 [ ] 2 [ ] 1 [ ] 0

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): __________ N

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

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

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. □ Indicated 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:

\[ \frac{\sum X}{N} = \]
\( X \) = individual quality of baseline ratings for each participant
\( N \) = number of participants in the study

Overall Rating for Quality of Baseline: (select 0, 1, 2, or 3): \[ \square 3 \quad \square 2 \quad \square 1 \quad \square 0 \]

(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).
**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</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Participant 1:</td>
<td>Behavior</td>
<td>Self Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Child</td>
<td>Teacher</td>
<td>Parent/sign. adult</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

|                       | Participant 2: | Behavior | Self Report | | |
|                       | Child | Teacher | Parent/sign. adult | | |
|                       | Ecology | Unknown | | | |
| Primary | | | | | |
| Secondary | | | | | |
| Unknown | | | | | |

|                       | Participant 3: | Behavior | Self Report | | |
|                       | Child | Teacher | Parent/sign. adult | | |
|                       | Ecology | Unknown | | | |
| Primary | | | | | |
| Secondary | | | | | |
| Unknown | | | | | |

|                       | Participant 4: | Behavior | Self Report | | |
|                       | Child | Teacher | Parent/sign. adult | | |
|                       | Ecology | Unknown | | | |
| Primary | | | | | |
| Secondary | | | | | |
| Unknown | | | | | |
### 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</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Participant 1:</td>
<td>Behavior</td>
<td>Self Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Child</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parent/sign. adult</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Participant 2:</td>
<td>Behavior</td>
<td>Self Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Child</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parent/sign. adult</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Participant 3:</td>
<td>Behavior</td>
<td>Self Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Child</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parent/sign. adult</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Participant 4:</td>
<td>Behavior</td>
<td>Self Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Child</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parent/sign. adult</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

### 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</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Participant 1:</td>
<td>Behavior</td>
<td>Self Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Child</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parent/sign. adult</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Participant 2:</td>
<td>Behavior</td>
<td>Self Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Child</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parent/sign. adult</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
<td></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</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>Participant 1:</td>
<td>Behavior</td>
<td>Self Report</td>
<td>Parent Report</td>
<td>Teacher Report</td>
</tr>
<tr>
<td>Secondary</td>
<td>Child</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parent/sign. adult</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>Participant 2:</td>
<td>Behavior</td>
<td>Self Report</td>
<td>Parent Report</td>
<td>Teacher Report</td>
</tr>
<tr>
<td>Secondary</td>
<td>Child</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parent/sign. adult</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>Participant 3:</td>
<td>Behavior</td>
<td>Self Report</td>
<td>Parent Report</td>
<td>Teacher Report</td>
</tr>
<tr>
<td>Secondary</td>
<td>Child</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parent/sign. adult</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>Participant 4:</td>
<td>Behavior</td>
<td>Self Report</td>
<td>Parent Report</td>
<td>Teacher Report</td>
</tr>
<tr>
<td>Secondary</td>
<td>Child</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parent/sign. adult</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ecology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>Secondary</td>
<td>Unknown</td>
<td>Child</td>
<td>Teacher</td>
<td>Parent/sign. adult</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
<td>---------</td>
<td>-------</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>Participant 1</td>
<td></td>
<td>(Answer C1.1 through C1.5)</td>
</tr>
<tr>
<td></td>
<td>Primary</td>
<td>C1.1 Change in levels (select one of the following)</td>
</tr>
<tr>
<td></td>
<td>Secondary</td>
<td>C1.1.1 large</td>
</tr>
<tr>
<td></td>
<td>Unknown</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 1: (select 0, 1, 2, or 3) □0 □1 □2 □3</td>
</tr>
<tr>
<td>Participant 2</td>
<td></td>
<td>(Answer C1.1 through C1.5)</td>
</tr>
<tr>
<td></td>
<td>Primary</td>
<td>C1.1 Change in levels (select one of the following)</td>
</tr>
<tr>
<td></td>
<td>Secondary</td>
<td>C1.1.1 large</td>
</tr>
<tr>
<td></td>
<td>Unknown</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 2: (select 0, 1, 2, or 3) □0 □1 □2 □3</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></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></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></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></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>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Approach Used</td>
</tr>
<tr>
<td>Participant 1</td>
<td>□ Primary □ Secondary □ Unknown</td>
<td>□ Baseline vs. intervention (e.g., ( A_1 + A_2 ) vs. ( B_1 + B_2 )) □ Baseline vs. follow up □ Unknown</td>
<td>□ Means and SDs □ Frequencies or proportions □ Other (specify): □ Unknown</td>
<td>ES = ______ □ Approach 1 □ Approach 2 □ Approach 3 □ Other</td>
</tr>
<tr>
<td>Participant 2</td>
<td>□ Primary □ Secondary □ Unknown</td>
<td>□ Baseline vs. intervention (e.g., ( A_1 + A_2 ) vs. ( B_1 + B_2 )) □ Baseline vs. follow up □ Unknown</td>
<td>□ Means and SDs □ Frequencies or proportions □ Other (specify): □ Unknown</td>
<td>ES = ______ □ Approach 1 □ Approach 2 □ Approach 3 □ Other</td>
</tr>
<tr>
<td>Participant 3</td>
<td>□ Primary □ Secondary □ Unknown</td>
<td>□ Baseline vs. intervention (e.g., ( A_1 + A_2 ) vs. ( B_1 + B_2 )) □ Baseline vs. follow up □ Unknown</td>
<td>□ Means and SDs □ Frequencies or proportions □ Other (specify): □ Unknown</td>
<td>ES = ______ □ Approach 1 □ Approach 2 □ Approach 3 □ Other</td>
</tr>
<tr>
<td>Participant 4</td>
<td>□ Primary □ Secondary □ Unknown</td>
<td>□ Baseline vs. intervention (e.g., ( A_1 + A_2 ) vs. ( B_1 + B_2 )) □ Baseline vs. follow up □ Unknown</td>
<td>□ Means and SDs □ Frequencies or proportions □ Other (specify): □ Unknown</td>
<td>ES = ______ □ Approach 1 □ Approach 2 □ Approach 3 □ Other</td>
</tr>
</tbody>
</table>
C3. Measures Support Primary Outcomes

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

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

\( X \) = 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:

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

\( X \) = 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><strong>D1. Categorical Diagnosis Data</strong></td>
<td>Diagnostic information regarding inclusion into the study presented: Yes ☐ No ☐ Unknown</td>
<td>Positive change in diagnostic criteria from baseline to intervention: Yes ☐ No ☐ Unknown</td>
<td>Positive change in diagnostic criteria from intervention to follow up: Yes ☐ No ☐ Unknown</td>
</tr>
<tr>
<td><strong>D2. Outcome Assessed via continuous Variables</strong></td>
<td></td>
<td>Positive change in number of participants showing clinical improvement from baseline to intervention: Yes ☐ No ☐ Unknown</td>
<td>Positive change in number of participants showing clinical improvement from intervention to follow up: Yes ☐ No ☐ Unknown</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: Yes ☐ No ☐ Unknown</td>
<td>Importance of behavior change from baseline to intervention is evaluated positively by individuals in direct contact with the participant: Yes ☐ No ☐ Unknown</td>
<td>Importance of behavior change from intervention to follow up is evaluated positively by individuals in direct contact with the participant: Yes ☐ No ☐ Unknown</td>
</tr>
<tr>
<td><strong>D4. Social Comparison:</strong> 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: Yes ☐ No ☐ Unknown</td>
<td>Participant’s behavior has improved from baseline to treatment when compared to normative data: Yes ☐ No ☐ Unknown</td>
<td>Participant’s behavior has improved from intervention to follow up when compared to normative data: 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

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>
</tbody>
</table>
### A3. Details are provided regarding variables that:

**A3.1** Have differential relevance for intended outcomes  

Specify: ____________________________

**A3.2** Have relevance to inclusion criteria  

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></td>
<td>Participants reported benefiting overall from the intervention</td>
</tr>
<tr>
<td>Parent/caregiver</td>
<td></td>
<td>Participants reported not benefiting overall from the intervention</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>
<tr>
<td>Child/Student</td>
<td></td>
<td>Participants reported benefiting overall from the intervention</td>
</tr>
<tr>
<td>Parent/caregiver</td>
<td></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  ______ N

   B2.2 months ______ N

   B2.3 years  ______ N

   B2.4 other  ______ N

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 ______ N

C2.2 frequency of intervention session ______ N

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</td>
</tr>
<tr>
<td></td>
<td>or 0 - 3</td>
<td>Promising</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weak</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No/limited evidence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or descriptive ratings</td>
</tr>
</tbody>
</table>

**General Characteristics**

- General Design Characteristics
- General Characteristics and Cultural Factors
- 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><strong>Descriptive or Supplemental Criteria</strong></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>
References


Kazdin, A. E. (1977). Assessing the clinical or applied significance of behavior change through social


APPENDIXES
APPENDIX A

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

Task Force, Co-Chairs

Thomas R. Kratochwill
Kimberly Eaton Hoagwood

Conceptual and Methodological Issues

Lynn Fuchs
Jan Hughes
Tim Keith
Joel Levin
Arthur Reynolds
Sylvia Rosenfield

Task-Wide and Classroom-Based Programs

Craig Frisby
Charles Maher
William Strein

Quantitative Methods Committee

Bonnie Nastasi
Steve Schensul
Doug Campbell
Denise DeZolt
David Fetterman
Karen Harris
Evelyn Jacob
Colette Ingraham
Margaret LeCompte
Joel Meyers
Evelyn Oka
Mittle Quinn
Jean J. Schensul
Cherie Tyler
Stephen Truscott
Kristen Varjas
Harry Wolcott
Frank Worrell

Multicultural Issues

Colette L. Ingraham
Evelyn Oka
Bonnie Nastasi

School-Based Prevention Programs

George Bear
Joe Durlak

School-Based Intervention Programs for Social Behavior Problems

Shane Jimerson
Greg Waas

Academic Intervention Programs

Virginia Berninger
Ed Shapiro

Family Intervention Programs

Cindy Carlson
Sandy Christenson

Comprehensive School Health Care

Rune Simeonsson
Ronda Talley

Research-to-Practice Committee

Susan Forman
Sandra Thompson

*Note: All members are listed in alphabetical order within domains/groups.
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 PSYCINFO 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 databases.

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; Gillham, 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 PSYC INFO 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 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:
  
  o To find articles about preventing depression in children and adolescents, we first start by typing in “depression” as the keyword.
  
  o 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.
  
  o We apply limits for the English language and childhood and adolescent age groups as we did previously, and now the results number 6,998.
  
  o 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.
  
  o 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.
  
  o 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.
  
  o 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.
  
  o 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 on Evidence-Based Interventions

General Resources on Conducting Literature Reviews

General Resources on Group and Single-Participant Design Meta-Analysis
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
  \[ X \ O_1 \]

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

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

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

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

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

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

- Untreated Control Group Design with Dependent Pretest and Posttest Samples

\[ \text{NR} \quad O_1 \times O_2 \]

\[ \text{NR} \quad O_1 \times O_2 \]

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

\[ \text{NR} \quad O_1 \quad O_2 \times O_3 \]

\[ \text{NR} \quad O_1 \quad O_2 \times O_3 \]

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

\[ \text{NR} \quad O_1 \times O_2 \quad O_3 \]

\[ \text{NR} \quad O_1 \times O_2 \times O_3 \]

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

\[ \text{NR} \quad O_1 \times \overset{-}{O_2} \]

\[ \text{NR} \quad O_1 \times \overset{+}{O_2} \]

- Cohort Control Group Design

\[ \text{NR} \quad O_1 \]

\[ \text{NR} \quad X \times O_2 \]

- Cohort Control Group Design with Pretest from Each Cohort

\[ \text{NR} \quad O_1 \quad O_2 \]

\[ \text{NR} \quad O_3 \times O_4 \]