

**Abstract Title Page**  
*Not included in page count.*

**Title:** Methods of Quality Appraisal for Studies Reviewed by Evidence Clearinghouses

**Authors and Affiliations:**

**Sandra Jo Wilson**

**Emily Tanner-Smith**

## **Abstract Body**

*Limit 4 pages single-spaced.*

### **Background and Objectives**

The second presentation will discuss quality appraisal methods for assessing research studies used in systematic reviews, research syntheses, and evidence-based practice repositories such as the What Works Clearinghouse. The different ways that the methodological rigor and risk of bias of primary studies included in syntheses is assessed means that different studies with greater or lesser quality might be included in the recommendations generated from such synthesis. Using the postsecondary education literature as an example, the presenters will describe how different methods of quality appraisal can result in potentially different reports on the extent of evidence on a topic and the level of confidence readers should have about that evidence.

Most evidence-based repositories, including the What Works Clearinghouse produce evidence reports for particular interventions. For example, the WWC has produced reports on interventions to teach mathematics (e.g., *Pre-K Mathematics*, *DreamBox Learning*) and interventions to prevent high school dropout (e.g., *Middle College High Schools*). CrimeSolutions.gov has reviewed a number of school-based interventions including career academies. These interventions are rather narrowly defined, and are often (though not always) a single branded program. This sort of evidence may be helpful for decision-making in two ways. Organizations already using a particular program may use such evidence to justify their choice of adopting the program after the fact. Or, should the evidence not be positive, an organization may use the evidence to back-up a decision to drop a program. In addition, organizations seeking new programming may use information on a number of different programs or strategies from an evidence clearinghouse to select the potentially most effective option. In either case, these are high stakes decisions. In thinking about how organizations might use evidence, therefore, it is not only a positive or a negative or significant effect that might be important. Other important issues that can weigh into decision-making have to do with the confidence we can place in the evidence; that is, with internal and external validity. How confident are we that a program is effective for producing the intended outcomes? How likely is it that a new organization will achieve the same results if they implement the same program?

This second paper will summarize the quality appraisal methods employed by several national research clearinghouses that produce evidence reports relevant to education, highlighting the elements of the quality appraisal and how the quality information is translated into the evidence recommendations that are reported to the public.

### **Population**

Several research clearinghouses provide evidence relevant to education. Foremost among these is the What Works Clearinghouse. According to the WWC's *Standards and Procedures Handbook (3.0)*, intervention reports "summarize all studies published during a specific time period that examine the effectiveness of an intervention" (p. 2). The conduct of intervention reports is governed by the *Handbook*, supplemented by topic specific protocols written with significant input from content experts. Together, these documents address eligible study designs, the required sample composition, specific nature of the intervention, the eligible outcomes, and how

eligible studies are graded. Relevant studies for the intervention reports are found through exhaustive searches of both the published and the unpublished literature.

In addition, nine other evidence-based repositories have been identified. The general quality appraisal procedures for each are described in Tables 1 and 2 below.

### **Research Design**

After describing the evidence-based practice repositories, the focus of the paper will be on describing and evaluating the quality appraisal procedures employed by the repositories. We will discuss how the repositories evaluate the internal and external validity of the candidate studies, and how that information is combined with the study findings to produce study ratings. The point here is not so much to emphasize the problems with the various methods, but more to promote understanding of how such quality appraisals can be interpreted.

The elements of the quality appraisal that will be discussed include:

1. Eligible outcome domains and the reliability and validity of outcome measures
2. Inclusion (or exclusion) of negative or harmful effects
3. Research design
4. Baseline equivalence
5. Attrition
6. The role of program developers
7. Implementation and cost information

Further details about the various evidence-based repositories are shown in Table 2.

### **Findings and Conclusions**

The presentation will conclude with a discussion of the implications of What Works Clearinghouse procedures on the kinds of evidence that are available to decisionmakers and the quality of that evidence.

## Appendix B. Tables and Figures

Not included in page count.

	Blueprints	California Evidence-Based Clearinghouse	Coalition for Evidence-Based Policy	Promising Practices Network	What Works in ReEntry
<b>Initial Outcome Screening</b>					
Includes outcomes <u>only</u> with favorable, significant impacts?	Yes	No	No	Yes	No
Includes outcomes with null or harmful effects?	No	Yes	Yes	No	Yes
Location of study must be U.S. or U.S. Territory?	Yes	Yes	Yes	Yes	Yes
<b>Review of Study Quality</b>					
<b>Confounding Factors Affecting Study Quality Ratings</b>					
Ratings or scores higher if baseline equivalent between conditions?	Yes	Not specified	Yes	Yes	Yes
Ratings or scores higher if overall or differential attrition rate is low?	Yes	Not specified	Yes	Not specified	Yes
Ratings or scores higher if overall or differential attrition bias is minimal?	Yes	Not specified	Yes	Not specified	Yes
Ratings or scores higher if ITT analysis used?	Yes	Not specified	Yes	Not specified	Not specified
Ratings or scores higher if analysis controls for baseline outcome measures (if applicable)?	Yes	Not specified	Yes	Not specified	Yes
Ratings or scores higher if analysis controls for age and gender?	Yes	Not specified	Yes	Not specified	Not specified
<b>Other Factors Affecting Study Quality Ratings</b>					
Ratings or scores higher if some way to measure program fidelity is used?	Yes	Yes	No	No	Yes
Ratings or scores higher if an independent evaluator is used?	No	Not specified	No	No	Yes
Ratings or scores higher if there is a larger sample size?	No (N's are required at each stage, however)	Not specified	Yes	Yes (requires n>30 in both groups)	Yes (requires n=30 in both groups for Basic; requires n=100 for both groups for High)
Ratings or scores higher if measures are reliable and valid?	Yes	Yes	Yes	No	Yes
<b>Conceptual Framework/Intervention Specificity Factors Affecting Study Quality Ratings</b>					
Ratings or scores higher if there is a theoretical foundation	No	No	No	No	No

	Blueprints	California Evidence-Based Clearinghouse	Coalition for Evidence-Based Policy	Promising Practices Network	What Works in ReEntry
Ratings or scores higher if prior research?	No	No	No	No	No
Ratings or scores higher if there is a clearly delineated program description?	Yes	Yes	No	No	No
<b>Review of Program Effectiveness</b>					
Comparative effectiveness studies (with no control) allowed?	No	No	No (only multiple RCTs allowed)	No	No
Subgroup Findings Reported? Are they Reviewed?	Yes	No	Not mentioned	Not specified	Yes
Uses 2-tailed significance test and <i>p</i> -value of less than .05 to determine significance?	Not specified	Not specified	Not specified	Yes, $p < 0.05$ ; Does not specify one- or two-tailed test	Yes, $p < 0.05$ ; Does not specify one- or two-tailed test
<b>Study Quality Classification</b>					
Is study quality characterized separately? How is it characterized?	Yes (Model program, Promising program)	Yes (Well Supported by Research Evidence, Supported by Research Evidence, Promising Research Evidence, Evidence Fails to Demonstrate Effect, Concerning Practice)	Yes (Top Tier, Near Top Tier)	Yes (Proven or Promising program)	Yes (High or Basic)
<b>Program Effectiveness Rating</b>					
Is there an effectiveness rating? What scale is used to describe effectiveness?	Not rated	Yes (Well Supported by Research Evidence, Supported by Research Evidence, Promising Research Evidence, Evidence Fails to Demonstrate Effect, Concerning Practice)	Not Rated	Not rated	Yes (Strong Evidence of Beneficial Effect, Modest Evidence of Beneficial Effect, No Statistically Significant Findings, Modest Evidence of Harmful Effect, Strong Evidence of Harmful Effect)
<b>* Role of Program Developer* (not part of review process)</b>					
Does program developer authorize posting of program summary to website?	Yes	Not specified	No	Not specified	Not specified
Does program developer select and prioritize studies and outcomes to be reviewed?	No	Not specified	No	Not specified	Not specified

	<b>SAMHSA NREPP<sup>§</sup></b>	<b>CrimeSolutions.gov/MPG/ FindYouthInfo.gov</b>	<b>IES What Works Clearinghouse</b>	<b>ASPE/OAH Teen Pregnancy Prevention</b>	<b>ACF Home Visiting Evidence of Effectiveness (HomeVee)</b>
<b>Initial Outcome Screening</b>					
<b>Includes outcomes <u>only</u> with favorable, significant impacts?</b>	<b>Yes</b>	No	No	No	No
<b>Includes outcomes with null or harmful effects?</b>	<b>No</b>	Yes	Yes	Yes	Yes
Location of study must be U.S. or U.S. Territory?	No Can be either if in English	No Can be either if in English	Yes Must be U.S. or U.S. Territory	Yes Must be U.S.-based	No Can be other developed- world context
<b>Review of Study Quality</b>					
<b>Confounding Factors Affecting Study Quality Ratings</b>					
Ratings or scores higher if baseline equivalent between conditions?	Yes	Yes	Yes	Yes Required for QED studies to be rated as moderate quality	Yes
Ratings or scores higher if overall or differential attrition rate is low?	Yes	Yes	Yes	Yes	Yes
Ratings or scores higher if overall or differential attrition bias is minimal?	Yes	Yes	Yes	Yes	Yes
Ratings or scores higher if ITT analysis used?	Yes	Yes	Yes	Yes	Yes
Ratings or scores higher if analysis controls for baseline outcome measures (if applicable)?	Yes	Yes	Yes	N/A Required for moderate rating (and for high rating if baseline nonequivalent)	Yes
Ratings or scores higher if analysis controls for age and gender?	Yes	Yes	Yes	N/A Required for high or moderate rating	Yes
<b>Other Factors Affecting Study Quality Ratings</b>					
Ratings or scores higher if some way to measure program fidelity is used?	Yes	Yes	No	N/A	N/A
Ratings or scores higher if an independent evaluator is used?	No	Yes	No	No	Yes
Ratings or scores higher if there is a larger sample size?	Yes	Yes	Yes	Yes	Yes
Ratings or scores higher if measures are reliable and valid?	Yes	Yes	Yes	Yes	Yes
<b>Conceptual Framework/Intervention Specificity Factors Affecting Study Quality Ratings</b>					
<b>Ratings or scores higher if there is a</b>	<b>No</b>	Yes	N/A	N/A	N/A

§ Key differences between NREPP and other evidence-based repositories are highlighted in **bold** font.

	SAMHSA NREPP <sup>5</sup>	CrimeSolutions.gov/MPG/ FindYouthInfo.gov	IES What Works Clearinghouse	ASPE/OAH Teen Pregnancy Prevention	ACF Home Visiting Evidence of Effectiveness (HomeVee)
<b>theoretical foundation</b>					
Ratings or scores higher if prior research?	No	Yes	N/A	N/A	N/A
Ratings or scores higher if there is a clearly delineated program description?	No	Yes	N/A	N/A	N/A
<b>Review of Program Effectiveness</b>					
<b>Comparative effectiveness studies (with no control) allowed?</b>	Yes	Yes But it depends on topic area	No	No	No
<b>Subgroup Findings Reported? Are they Reviewed?</b>	Yes/No Reported but not reviewed	Yes/Yes Reported and reviewed, but on a case-by-case basis	Yes/No WWC presents the subgroup results as supplemental tables. Separate subgroup results do not average into the intervention rating	Yes/Yes TPP reports and reviews subgroup findings for gender and sexual experience subgroup	Yes/Yes HomeVee reports and reviews subgroup findings if such findings are replicated in the same outcome domain in at least two studies using different analytic samples
<b>Uses 2-tailed significance test and <math>p</math>-value of less than .05 to determine significance?</b>	No NREPP considers outcomes evaluated using a 1- or 2-tailed significance test and an alpha level equal to .05 significant	No ( $p$ value can be $p = .05$ )	Yes	Yes	Yes
<b>Study Quality Classification</b>					
<b>Is study quality characterized separately? How is it characterized?</b>	Yes Numeric score (0-4)	No Rating contributes to program effectiveness rating	Yes Meets standards without reservations, meets standards With reservations, does not meet standards	Yes High, Moderate, Low	Yes High, Moderate, Low
<b>Program Effectiveness Rating</b>					
<b>Is there an effectiveness rating? What scale is used to describe effectiveness?</b>	No NREPP does not rate program or outcome effectiveness	Yes Effective, Promising, No Effects (program-level)	Yes Positive, Potentially Positive, Mixed, Indiscernible, Potentially Negative, Negative (outcome-level)	Not rated	Not rated
<b>* Role of Program Developer* (not part of review process)</b>					
<b>Does program developer authorize posting of program summary to website?</b>	Yes	No	No	No	No

	SAMHSA NREPP <sup>5</sup>	CrimeSolutions.gov/MPG/ FindYouthInfo.gov	IES What Works Clearinghouse	ASPE/OAH Teen Pregnancy Prevention	ACF Home Visiting Evidence of Effectiveness (HomeVee)
Does program developer select and prioritize studies and outcomes to be reviewed?	Yes	No	No	No	No