

<b>Symposium Title</b>	<i>New Directions for the What Works Clearinghouse (WWC): Expanding the Accessibility and Reach of WWC Products</i>
<b>Symposium Contact</b>	Allan Porowski ( <a href="mailto:allan_porowski@abtassoc.com">allan_porowski@abtassoc.com</a> )
<b>Preferred Conference Sections</b>	Research Methods, Effects of Education Policies
<b>Symposium Justification and Abstracts</b>	See below
<b>Names of Papers</b>	<p><i>Prioritization of Topic Areas and Interventions in Early Childhood to Grade 12 for the What Works Clearinghouse</i></p> <p><u>Authors:</u>  Elias Walsh (<b>presenting author and contact</b>, Mathematica Policy Research, <a href="mailto:EWalsh@mathematica-mpr.com">EWalsh@mathematica-mpr.com</a>)  Jill Constantine (Mathematica Policy Research, <a href="mailto:JConstantine@mathematica-mpr.com">JConstantine@mathematica-mpr.com</a>)</p> <p><i>New approaches for characterizing findings in What Works Clearinghouse Intervention Reports</i></p> <p><u>Authors:</u>  Joshua Polanin (<b>presenting author and contact</b>, American Institutes for Research, <a href="mailto:jpolanin@air.org">jpolanin@air.org</a>)  Jeff Valentine (University of Louisville, <a href="mailto:jeff.valentine@louisville.edu">jeff.valentine@louisville.edu</a>)  Ryan Williams (American Institutes for Research, <a href="mailto:rwilliams@air.org">rwilliams@air.org</a>)  Joe Taylor (American Institutes for Research, <a href="mailto:jotaylor@air.org">jotaylor@air.org</a>)  Sarah Caverly (American Institutes for Research, <a href="mailto:scaverly@air.org">scaverly@air.org</a>)  Daniel Hubbard (American Institutes for Research, <a href="mailto:dhubbard@air.org">dhubbard@air.org</a>)  Elizabeth Nolan (American Institutes for Research, <a href="mailto:enolan@air.org">enolan@air.org</a>)</p> <p><i>Expanding the Accessibility and Reach of Systematic Evidence Reviews through New Reporting Formats</i></p> <p><u>Authors:</u>  Allan Porowski (<b>presenting author and contact</b>, Abt Associates, <a href="mailto:allan_porowski@abtassoc.com">allan_porowski@abtassoc.com</a>)  Michael Frye (Abt Associates, <a href="mailto:michael_frye@abtassoc.com">michael_frye@abtassoc.com</a>)  Sandra Wilson (Abt Associates, <a href="mailto:sandra_wilson@abtassoc.com">sandra_wilson@abtassoc.com</a>)</p>

<b>Independent Discussant (confirmed)</b>	Christopher Weiss (U.S. Department of Education, Institute of Education Sciences, <a href="mailto:Christopher.Weiss@ed.gov">Christopher.Weiss@ed.gov</a> )
---	--

### **Symposium Justification**

The What Works Clearinghouse (WWC) was established in 2002 to serve as a central and trusted resource for evidence on the effectiveness of educational interventions. Since then, the WWC’s systematic evidence review methods and presentation formats have been adopted, at least in part, by numerous federal and state agencies, as well as by evidence collaboratives in the private sector. In the past year, the WWC underwent a reorganization of its pK-12 topic area reviews, and work continued in the postsecondary topic areas under a separate contract. In this symposium, presenters from the three primary WWC contracts will share new directions for organizing, summarizing results, and presenting findings from systematic evidence reviews for the next generation of the WWC.

The first presentation will be delivered by the contractor for the What Works Clearinghouse Organization of Reporting of Evidence on Graduation, Achievement, and Nonacademic Outcomes (OREGANO) project, which guides the WWC review process. This presentation will focus on new methods to select topics for review by reconciling the supply of evidence available, indicators of consumer demand for evidence, and guidance from content experts. This presentation will also include an overview of the new topics selected for review in 2018-19.

The second presentation will be presented by the contractor for the What Works Clearinghouse: Statistics, Website, and Training (SWAT) project, which guides the development of WWC evidence standards and procedures. This presentation will focus on new methods to summarize the evidence from systematic reviews that are technically sound yet accessible to a lay audience. In particular, this presentation will include a discussion of how findings from multiple studies can be synthesized and summarized in ways that provide greater precision and simplicity for public consumption.

The final presentation will be presented by the contractor of the What Works Clearinghouse: Postsecondary Education, Postsecondary Preparation, and Evidence Reporting (WWC-PEPPER) contract, which guides reviews for postsecondary topic areas. In this presentation, the WWC-PEPPER team will summarize efforts to modify the What Works Clearinghouse intervention report format so that evidence is accessible. The presentation will highlight new design elements of the intervention report, including the inclusion of purpose statements, plain-language summaries of ratings, detailed descriptions of intervention components, systematic reporting of cost data, context infographics, and new templates for findings tables. This presentation will not only provide an update on the WWC’s product lines, it also will provide researchers with ideas for summarizing results and connecting with practitioner audiences.

Dr. Christopher Weiss will serve as the independent discussant for this symposium.

## ***Prioritization of Topic Areas and Interventions in Early Childhood to Grade 12 for the What Works Clearinghouse***

**Elias Walsh and Jill Constantine (Mathematica Policy Research)**

### **Background:**

Before assessing the quality of research on education interventions, and summarizing and disseminating findings from its systematic reviews, the What Works Clearinghouse (WWC) must identify and prioritize topics and interventions for review. This process must balance the limited resources available for conducting reviews with the goal of comprehensive coverage of the relevant literature.

The WWC: Organization of Reporting of Evidence on Graduation, Achievement, and Nonacademic Outcomes (WWC-OREGANO) team will produce intervention reports in a variety of topic areas in early childhood, elementary, and secondary education. Promising topics for these reports include those that have sufficient evidence to support an *informative* report and those that are *relevant* to decision makers. Once topics are identified, topic area teams develop review protocols that describe how reviewers will conduct systematic reviews of interventions in each topic area, including a description of what studies are eligible for review, how the WWC will search for them, and how they will be reviewed. To ensure that intervention reports reflect content area expertise, each topic area team tailors review protocols based on input from prominent experts in the field.

### **Objective:**

In this presentation, we will (1) describe a process for selecting relevant and informative early childhood to grade 12 topic areas and interventions for systematic WWC reviews, and (2) describe a process for developing protocols to guide the WWC reviews in these topic areas. Specifically, we will present the process WWC-OREGANO used to determine topic areas for reviews, and we will also outline the process for developing review protocols, their key elements, and how content experts are consulted and how their expertise informs the development of the protocols.

### **Approach:**

We use a two-step process to identify the targets for WWC-OREGANO reviews. First, we use a framework to select topic areas that considers both the *demand* for evidence (whether the topic is relevant to decision makers) and the *supply* of evidence (whether it can support informative products). We then develop lists of interventions in each subject area, which we then research to determine the extent of available rigorous evidence. We include interventions that were (1) recommended by the content experts; (2) the subject of older intervention reports that remain relevant to decision makers but were developed under previous versions of the WWC standards; or (3) evaluated by individual studies reviewed by the WWC but not included in an intervention report. Finally, to support reviews of studies for these interventions, we develop review protocols

that describe what studies are eligible for review, how the WWC will search for them, and how they will be reviewed.

**Conclusion:**

This prioritization process is designed to identify topics for WWC intervention reports that best meet the needs of education decision makers. This presentation will provide an opportunity to share this process with researchers and other stakeholders and obtain their feedback to improve it in the future.

## *New approaches for characterizing findings in What Works Clearinghouse Intervention Reports*

**Jeff Valentine (University of Louisville), Ryan Williams, Joe Taylor, Sarah Caverly, Daniel Hubbard, Elizabeth Nolan, & Joshua R. Polanin (American Institutes for Research)**

### **Background/Context:**

One of the many purposes of the What Works Clearinghouse (WWC) is to evaluate the effectiveness of a single intervention or program, such as a curriculum or teacher training program, by evaluating and synthesizing the evidence presented in all studies that meet WWC's standards. To date, this has been done with a vote-counting method; the relative numbers of studies finding positive, negative, and null effects are summarized, with each study with the same design rating counting equally without regard for its effect size, sample size, design, or representativeness<sup>1</sup>. For instance, if two studies with small sample sizes found null effects while a third study of the same intervention, with a large sample size, found a strong positive effect, an intervention report for the three studies would characterize the results of the intervention as "mixed effects". Particularly in small samples, the current protocol errs on the side of conservatism but does so in an imprecise way; as the number of studies grows, the standards move from being overly conservative to overly liberal.

As such, we aim to improve the procedures for synthesizing multiple studies of the same intervention, to increase precision and to ensure that the characterization accurately reflects the body of literature it intends to represent.

### **Practice:**

We propose two potential methods for synthesizing the results of group design studies. Further, we propose to use the results of the synthesis model as a basis for characterizing findings from intervention reports.

The first potential method for synthesizing results is an unweighted effect size with a confidence interval. This method creates a summary effect of the full body of research and drawing inferences about study quality from its attributes. Summary effects with larger magnitudes are deemed stronger; summary effects with tighter confidence intervals are deemed more trustworthy. Alternatively, a second potential method is to use meta-analysis methods with fixed or random effects, frequently used in other fields such as the health sciences to distill several experiments into a single effect size. These would incorporate various weighting schemes to determine the contribution of each study to the overall estimate.

Because each of these methods allows us to combine multiple studies into a single effect size and confidence interval for a given intervention, we also propose an adjustment to the rating system used in intervention reports. Interventions that, when distilled into a single effect size and confidence interval, meet minimum levels of magnitude and statistical significance, can be characterized as Positive or Negative; those that meet one of those criteria but not the other can be called Potentially Positive or Negative (assuming the confidence interval is not unreasonably

---

<sup>1</sup> See What Works Clearinghouse Procedures Handbook v4.0, or Table 1 of this document.

wide); and those that have very wide confidence intervals, regardless of sign or magnitude, fall into No Discernible Effects. Table 2 provides a full explanation.

To examine how the various methods of combining study effects would influence the distribution of WWC effectiveness ratings, we compiled 116 meta-analyses of 62 interventions and assigned them to ratings of Positive, Potentially Positive, No Discernible Effects, Potentially Negative, and Negative based on the underlying data and our chosen method of determining study quality.<sup>2</sup> Compared against the baseline data from the unweighted averages, the weighted averages are almost universally more conservative. All of the weighted models assigned a Negative or Potentially Negative effect to only one treatment, while the unweighted model placed 12 in the former category and nine in the latter; similarly, all of the weighted models had more (often many more) treatments in the category of No Discernible Effects. The fixed-effects model assigned a nearly-equal number of treatments to Positive or Potentially Positive compared with the unweighted model, while the random-effects models were also somewhat more conservative in assigning studies to the Positive category. Results are shown in Table 3.

### **Discussion:**

Each design comes with its respective tradeoffs. The random-effects design is very cautious, viewing studies with a skeptical eye that reduces the number of false positives; it is also difficult to interpret and to fine-tune, with a variety of estimators and adjustments available, none of them used universally in the literature. In turn, the fixed-effects design is a bit more rigid and has fewer options (perhaps a mixed blessing), and its assumptions are rarely accurate in practice, but researchers can discern the impact that those assumptions have and assess the extent to which the model is inaccurate.

There are also interpretation concerns to keep in mind when choosing a method for synthesizing multiple studies into a single effect and rating. For one, it is important that the method produces variation in the results, otherwise the results lose meaning. While conservatism is important, if the criteria are so conservative that almost no interventions earn ratings other than No Discernible Effects, it becomes impossible to distinguish better interventions from worse ones. Additionally, while the summary should be concise, it should also provide at least as much information as the sum of its parts; it should take care not to smooth over the details of the individual studies in the name of creating a single coherent narrative.

We are in the process of determining the most appropriate model to use, given the circumstances of the review (i.e., the number of studies available to synthesize, the amount of between-study heterogeneity). While we are not ready to state publicly the exact details of the model, it will rely on the work conducted in this presentation and will be out for public consumption in late 2018. We expect that researchers relying on the results of Intervention Reports will find the approach is an advance in the way the WWC synthesizes studies. And ultimately, users of Intervention Reports will find the results illustrate clearly the syntheses' findings.

---

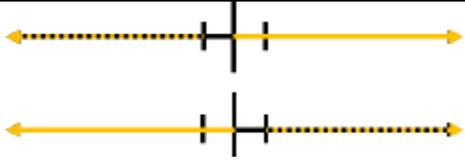
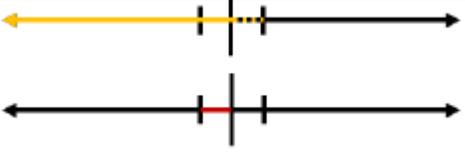
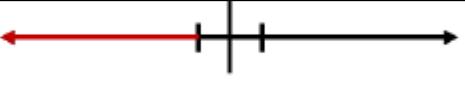
<sup>2</sup> Our criteria for assigning studies to these ratings are very similar, but not identical, to those used by WWC.

**Table 1: Characterization of Findings from Existing WWC Handbook**

<p>Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.</p>	<ul style="list-style-type: none"> <li>• Two or more studies show statistically significant positive effects, at least one of which meets WWC group design standards without reservations, AND</li> <li>• No studies show statistically significant or substantively important negative effects.</li> </ul>
<p>Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.</p>	<ul style="list-style-type: none"> <li>• At least one study shows statistically significant or substantively important positive effects, AND</li> <li>• Fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects, AND</li> <li>• No studies show statistically significant or substantively important negative effects.</li> </ul>
<p>No discernible effects: No affirmative evidence of effects.</p>	<ul style="list-style-type: none"> <li>• None of the studies show statistically significant or substantively important effects, either positive or negative.</li> </ul>
<p>Mixed effects: Evidence of inconsistent effects.</p>	<p>EITHER both of the following:</p> <ul style="list-style-type: none"> <li>• At least one study shows statistically significant or substantively important positive effects, AND</li> <li>• At least one study shows statistically significant or substantively important negative effects, BUT no more such studies than the number showing statistically significant or substantively important negative effects.</li> </ul> <p>OR both of the following:</p> <ul style="list-style-type: none"> <li>• At least one study shows statistically significant or substantively important effects, AND</li> <li>• More studies show an indeterminate effect than show statistically significant or substantively important effects.</li> </ul>
<p>Potentially negative effects: Evidence of a negative effect with no overriding contrary evidence.</p>	<p>EITHER both of the following:</p> <ul style="list-style-type: none"> <li>• One study shows statistically significant or substantively important negative effects, AND</li> <li>• No studies show statistically significant or substantively important positive effects.</li> </ul> <p>OR both of the following:</p> <ul style="list-style-type: none"> <li>• Two or more studies show statistically significant or substantively important negative effects, at least one study shows statistically significant or substantively important positive effects, AND</li> <li>• More studies show statistically significant or substantively important negative effects than show statistically significant or substantively important positive effects.</li> </ul>
<p>Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.</p>	<ul style="list-style-type: none"> <li>• Two or more studies show statistically significant negative effects, at least one of which meets WWC group design standards without reservations, AND</li> <li>• No studies show statistically significant or substantively important positive effects.</li> </ul>

Note: A statistically significant estimate of an effect is one for which the probability of observing an effect that is at least as large as the measured effect under the view that the intervention had no impact is less than one in 20 (using a two-tailed t-test with  $p = 0.05$ ). An effect size of 0.25 standard deviations or larger is considered to be substantively important. An indeterminate effect is one for which the single or mean effect is neither statistically significant or substantively important.

**Table 2: A Proposed Intervention Ratings Scheme**

Positive Effects	At least two studies have an average effect size of at least 0.05 standard deviations and statistically significant.	
Potentially Positive Effects	At least two studies have an average effect size is positive, not statistically significant, and the margin of error does not include a number smaller than -0.05, OR One study has an effect size that is positive, less than 0.05 standard deviations, and statistically significant.	
No Discernible Effects	At least two studies have an average effect size that is positive, not statistically significant, and includes a number smaller than -0.05 within the margin of error, OR One study has an effect size that is negative, not statistically significant, and includes a number larger than 0.05 within the margin of error.	
Potentially Negative Effects	At least two studies have an average effect size is negative, not statistically significant, and the margin of error does not include a number larger than 0.05, OR One study has an effect size that is negative, larger than -0.05 standard deviations, and statistically significant.	
Negative Effects	At least two studies have an average effect size of 0.05 standard deviations or smaller and statistically significant.	

A solid red line indicates a statistically significant effect. A solid yellow line indicates an effect size estimate that may not be statistically significant. A dashed yellow line indicates the acceptable range for the lower bound of the margin of error (confidence interval) for a positive effect, or an upper bound of the margin of error for a negative effect.

**Table 3: Differences in Effectiveness Ratings Across Multiple Approaches**

	Positive	Potentially Positive	No Discernible Effects	Potentially Negative	Negative
Using Unweighted Average and CI	49 (42.2%)	13 (11.2%)	33 (28.4%)	9 (0.7%)	12 (10.3%)
Using Fixed Effect Weighted Average and CI	50 (43.1%)	14 (12.1%)	51 (44.0%)	0 (0%)	1 (0.9%)
Using Random Effects Weighted Average and CI (REML)	37 (31.9%)	17 (14.7%)	61 (52.6%)	0 (0%)	1 (0.9%)
Using Random Effects Weighted Average and CI (REML with K-H)	15 (12.9%)	10 (8.6%)	90 (77.6%)	0 (0%)	1 (0.9%)
Using Random Effects Weighted Average and CI (DL)	37 (31.9%)	17 (14.7%)	61 (52.6%)	0 (0%)	1 (0.9%)
Using Random Effects Weighted Average and CI (DL with K-H)	15 (12.9%)	10 (8.6%)	90 (77.6%)	0 (0%)	1 (0.9%)

Approach based on 116 meta-analyses of 62 interventions and ignores the distinction between studies that meet evidence standards with and without reservations.

REML = restricted maximum-likelihood estimator; DL = DerSimonian-Laird estimator; K-H = Knapp-Hartung adjustment

## ***Expanding the Accessibility and Reach of Systematic Evidence Reviews through New Reporting Formats***

**Michael Frye, Sandra Wilson, and Allan Porowski (Abt Associates)**

### **Background/Context:**

The What Works Clearinghouse (WWC) was established in 2002 to serve as a central and trusted resource for evidence on the effectiveness of educational interventions. Coders certified in the application of WWC evidence standards review studies and then synthesize their findings for the public in intervention reports or practice guides. Intervention reports summarize evidence on a particular intervention and include a WWC effectiveness rating that incorporates elements of both the rigor of the research and the significance of the findings. Practice guides are developed by a panel of national experts – including researchers and practitioners – and draw upon the evidence to provide a list of recommendations for educators and policymakers.

Although these templates undergo periodic revisions, they have maintained the same basic look and feel for more than a decade. In mid-2018, the U.S. Department of Education’s Institute of Education Sciences (IES) signaled a desire to rethink intervention report templates. The guidance from IES indicated a desire to explore several new directions:

- Providing more detailed information on the cost of interventions
- Communicating findings more clearly to a range of stakeholders, by simplifying language, developing research metrics that are easier to understand
- Rethinking the reliance on statistical significance in describing findings
- Broadening coverage of the WWC’s offerings to include interventions that focus on development of children age birth to three, and expanding coverage in postsecondary education (and particularly with regard to career and technical training).<sup>3</sup>

These new directions are all motivated by an underlying goal – improving the way in which WWC communicates results to educators and policymakers.

### **Purpose/Objective/Research Question:**

In this session, authors describe their experiences redesigning intervention report templates to ensure their utility among diverse audiences. Our core research question is: *How can researchers more effectively communicate the research findings to non-researcher audiences?*

### **Research Design and Data Collection:**

The team developing the new intervention report templates first conferred with literature on plain-language presentation of research findings. Several sources were identified that validated innovative presentation strategies to summarize evidence (Carrasco-Labra et al., 2016<sup>4</sup>), that

---

<sup>3</sup> <https://ies.ed.gov/director/wwc2018.asp>

<sup>4</sup> Carrasco-Labra, A., Brignardello-Petersen, R., Santesso, N., Neumann, I., Mustafa, R.A. et al. (2016). Improving GRADE evidence tables part 1: A randomized trial show improved understanding of content in summary of findings tables with a new format. *Journal of Clinical Epidemiology*, 74, 7-18.

provided plain-language summary statements on the findings of a systematic review (Cochrane Norway, 2017<sup>5</sup>), and that summarized evidence for dichotomous outcomes (Hahn et al., 2015)<sup>6</sup>. The team also conferred with professional editors, professional designers and desktop publishers, and other researchers internally to identify the most promising ideas for the new templates.

## Results:

The team tasked to redesign the intervention report templates made several key improvements to the intervention report templates. These improvements are summarized here:

- **Inclusion of purpose statements:** IES provided specific guidance to include a purpose statement at the beginning of an intervention report. This purpose statement provides background on the core challenge being addressed by the intervention under review, and connects the intervention to the larger body of work reviewed in the topic area.
- **Plain-language summaries of findings:** The WWC addressed IES's desire to summarize findings without technical jargon by drawing up plain-language statements that summarize the evidence rating of the study and the intervention's effects on each eligible outcome.
- **Systematic descriptions of interventions:** The WWC intervention descriptions have been modified to provide systematic details on the goal, target population, method of delivery, frequency/duration of service, and intervention components. These standard descriptors provide a more consistent level of detail in the intervention reports and are designed to provide WWC consumers with the requisite information needed to assess whether implementation of a given intervention is feasible.
- **Systematic reporting of cost data:** The WWC has developed a systematic approach to describing the cost of interventions, using categories for personnel costs, facilities costs, equipment and materials costs, costs paid by students/parents, in-kind supports, and sources of funding. These categories allow consumers to identify cost drivers, resource needs, and opportunities for financial or in-kind support for implementing the intervention.
- **Context infographics:** While the WWC has an Extent of Evidence rating that summarizes the size of study sample, this section is designed to provide a deeper investigation of the context of the study so WWC consumers can assess the external validity of the research. This section includes a description of the participants in the research and sample characteristics across all studies summarized. This information can

---

<sup>5</sup> Cochrane Norway, 2017. How to write a plain language summary of a Cochrane intervention review: Checklist. 28 February. Author.

<sup>6</sup> Hahn, R.A., Knopf, J.A., Wilson, S.J., Truman, B.I., Milstein, B. et al. (2015). Programs to increase high school completion: A community guide systematic health equity review. *American Journal of Preventive Medicine*, 48(5), 599-608.

help educators understand the context in which the evidence was developed, and determine whether the program might be suitable for their setting.

- **Plain-language summary tables:** New table formats in the intervention report provide readers with individual study findings that can lead the reader on a narrative of the findings instead of presenting static tables of numbers. These table formats show the outcome, effectiveness rating, the performance of the treatment group relative to the comparison group, and the extent of evidence meeting WWC standards (i.e., the number of studies and students).

### **Conclusions:**

Although these intervention reports have not been finalized at the time of this writing, these new templates are expected to be ready for public release by March 2019. This presentation provides an opportunity for researchers to identify innovative ways of presenting findings that will reach audiences more effectively and connect research to practice in a more systematic and integrated fashion.