

# REVIEW PROTOCOL FOR DATA-DRIVEN DECISION MAKING PRACTICE GUIDE

## SCREENING CRITERIA

### Issue: **Time Frame**

Relevant to: Screening Criteria

Protocol:

The WWC reviews studies published in the last 20 years. Reviewers should contact the evidence coordinator about studies published before that time to determine whether they should be reviewed.

### Issue: **Location and Language**

Relevant to: Screening Criteria

Protocol:

All interventions must be conducted in English and must be conducted in the United States.

### Issue: **Age and/or Grade Range**

Relevant to: Screening Criteria

Protocol:

The target grade range for the data-driven decision making (DDDM) practice guide is **kindergarten through twelfth grade**, or ages 5 through 18. If you are assigned a study that is not in this range, please check with Cassie before reviewing. No specific populations will be targeted. However, reviewers should note in the SRG and on the study tracker summary section if a study only focused on a specific target population.

### Issue: **Study Design**

Relevant to: Screening Criteria

Protocol: **RCTS, QEDs that meet WWC standards, Single-Case Designs, and Regression Discontinuity Designs are eligible for full WWC review.**

**Correlational, observational, and qualitative studies have a different SRG that should be completed for reviews of these study designs.**

There is limited rigorous research examining whether alternative structures for data-driven decision making have different impacts on educators' practices or on student outcomes. The panel will examine those studies with rigorous research designs, as well as those which use correlational or qualitative data. The panel recognizes that recommendations based on correlational or qualitative research are likely to be classified as having low evidence, but these studies are still worth consideration and should be reviewed according to the correlational/qualitative SRG. The evidence coordinator, practice coordinator, or another designated WWC staff member will pre-screen studies before assigning them to reviewers to identify the study design, marking the design in the study tracker. Reviewers should confirm the study design when reviewing a study. Studies with comparison groups (RCTs and QEDs) will be assigned to certified reviewers. Qualitative studies with comparison groups (for example, those comparing DDDM implementation across several districts) will be summarized but will not be subject to a formal WWC review. Those without comparison groups (qualitative and correlational) will not meet standards, and may be assigned to either certified reviewers or designated WWC staff who will summarize the key

points of the study’s correlational or qualitative design and its findings, to assist the panel with making recommendations. The summaries of qualitative and correlational studies may be used at the panel’s discretion to support panel recommendations.

Regression Discontinuity Designs (RDDs) and Single-Case Designs (SCDs) are eligible for review. Until WWC RDD Standards are disseminated, please make note that the study is an RDD or SCD and will have to be sent for review by an expert in RDDs or SCDs.

**Issue: Brief Description of Intervention**

Relevant to: Screening Criteria

Protocol

This DDDM Practice Guide is developing recommendations on how to use student data to make decisions about **instructional practices**. Specifically, this practice guide is focused on how to alter instructional practices in a specific school or classroom **after the school year has commenced** (as opposed to changes between school years). The guide will inform administrators (principals as well as district staff) and educators how best to use assessment data on their students to determine whether and how to make changes to instruction (such as when to revisit concepts, accelerate lessons, provide supplemental resources, etc.). The guide will only provide recommendations for using **standardized assessment data** to make instructional decisions.

For the purpose of this guide, “standardized” assessments are those that are administered and scored consistently within a district, region, state, etc. They may be either formative (designed to inform practice as learning is taking place) or summative (designed to evaluate learning at the end of a unit of study), and their results may be reported in respect to an empirical distribution (“norm-referenced” reporting) or a target achievement level (“criterion-referenced” reporting). Recommendations and practices will likely focus on issues below through the use of standardized assessment data:

- Existing student assessment, data systems, and data-driven decision-making processes that have demonstrated effectiveness in improving student instruction and achievement. (Assessment Systems)
- Specific instructional practices related to data use that have demonstrated effectiveness in improving student achievement. (Instructional Practices)
- Design and implementation of district-wide data systems that ensure critical elements of student achievement are tracked. (Designing/Implementing Data Systems)
- Summarizing and reporting data to best inform decisions at the district, school, classroom and student levels. (Approaches to Analyzing/Summarizing/Reporting Data)
- Assisting administrators and instructional staff in synthesizing, interpreting, and using data to improve instructional decision making. (Leadership Practices; Collaborative Practices; Instructional Practices)

Recommendations will NOT address the following:

- Use of data from assessments that teachers develop and use only in their own classrooms.

- Use of data for the exclusive purpose of grouping students or determining when low performing students need supplemental instruction (e.g., use of data for determining special education or gifted student classification).
- Use of data to evaluate teacher and/or administrator performance.

Note that studies will be eligible if the intervention uses annual performance data as one of the pieces of data for making instructional decisions. For instance, schools may use annual performance data as a benchmark, and make decisions within the school and within the school year based on students' performance relative to those benchmarks.

Additionally, studies can be eligible if they use annual performance data (including VAM data) as an **outcome** in the study. That is, the use of VAM data to make decisions about curriculum or practices is within the scope of the practice guide. For example, a study may examine the effects of a decision making process implemented in some schools to inform instructional decisions throughout the school year, and it may assess the effectiveness by comparing intervention schools' VAM estimates with non-intervention schools' VAM estimates. What makes this example study eligible is that the schools in the intervention used data on their students from within the school year to make instructional decisions within the school year.

## FULL REVIEW

### Issue: **Outcomes and Domains for Multiple Comparisons**

Relevant to: Screening Criteria, Full Review

Protocol:

Eligible outcomes fall into two broad categories.

#### ***Proximal Outcomes***—Education Practices:

- *Leadership Practices*. Includes leadership steps taken by administrators and/or teachers.
- *Collaborative Practices*. Includes collaborative data use by teachers, administrators, or both.
- *Instructional Practices*. Includes teaching practices employed by teachers in light of their data use.
- *Communication with External Entities*. Demonstrates how schools and districts are sharing data, and aiding in data consumption and understanding, for external entities, such as parents and community organizations (e.g., growth reports based on percentiles).

Proximal outcomes may be discussed in the practice guides, but are typically studied with more qualitative designs and may contribute a lower level of evidence. Among these non-student outcomes, those that can be concretely defined (how, how often, how many) will be more useful to the panel than outcomes for which measurement accuracy is more difficult (fidelity, staff buy-in or reaction, school climate or atmosphere).

#### ***Intermediate Outcomes***—Student Achievement/Performance:

- Mathematics Performance
- Reading Performance
- Writing Performance
- Other Academic Performance
- Retention in Grade
- High School Graduation
- Pursuit of Postsecondary Education

NOTE: *Adjustments for multiple comparisons should be made when there are multiple measures of students' academic performance on standardized tests, as in a study that examines both reading and mathematics performance.*

Any claims of effectiveness for a DDDM intervention/practice will need to be supported by student achievement data in order for the intervention/practice to obtain a rating of moderate or high level of evidence.

When completing the study tracker after a review, reviewers should mark whether the study included any non-student outcomes in the field above the Outcomes and Effects by Outcome summary box.

Issue: **Criteria for QED Equivalence**

Relevant to: Full Review

Protocol:

Authors need to demonstrate equivalence on **student achievement** at pretest [if student achievement/performance is the outcome; however, if demographics (e.g., SES) are not balanced, please note this on the Study Review Guide].

Issue: **Study Ratings**

Relevant to: Full Review

Protocol:

Ratings include: *Meets standards*, *Meets standards with reservations*, *Does not meet standards*, and *Uncertain* (*uncertain* will be used in cases where an author query would be necessary to determine rating).

Issue: **Description of Intervention**

Relevant to: Full Review

Protocol:

Descriptions of Interventions should be detailed. Among relevant items, as indicated in “**Brief Description of Intervention: Screening Criteria**,” reviewers should record the following, when possible:

- **Intervention location/setting** (e.g., general education classroom, school, district, etc.).
- **Person who delivers the intervention** (e.g., general education teacher, resource assistant, administrator, etc.).
- **Duration and Frequency of data collection** (e.g., length of intervention/data collection period).
- **Type and Frequency of Professional Development** (e.g., training, coaching) related to data use
- **Format** in which student assessment results are available to educators (e.g., extent to which results are disaggregated by student characteristics and/or by academic content; also, extent to which results can be manipulated electronically for ease of analysis).
- **Design of Assessment System Details.** This could include assessment format (e.g., electronic or paper administration); content and format of the assessment items (e.g., whether it is explicitly aligned with the curriculum; whether items are exclusively multiple-choice); psychometric properties of the system; and the role of the system in the school or district accountability system (e.g., whether results are reported and used at the district level for accountability purposes).
- **Materials** (e.g., computer-based, software used, etc.)
- **Any features that distinguish this intervention** from others (e.g., teacher professional development included, use of visuals to convey data, etc.).

Issue: **Missing Information**

Relevant to: Full Review

Protocol:

Because of the tight timeline for practice guides, authors are typically not contacted as part of the review process. If the **study is missing information necessary to make a rating, rate as *uncertain* and note the information that is missing**. (If a study is critical for determining the level of evidence for a recommendation, then we may revisit it and contact the author.)

Studies exhibiting overall attrition greater than or equal to 20%, or differential attrition greater than or equal to 7%, will be downgraded. RCTs exhibiting this problem can receive a maximum rating of *meets standards with reservations*, and to do so, must demonstrate baseline equivalence of the analysis sample. QEDs are not subject to attrition standards.

QEDs (and RCTs with high attrition as defined above) that do not demonstrate baseline equivalence will fail to meet evidence standards. To demonstrate equivalence, the intervention and comparison groups must:

- Not be significantly different at baseline (after accounting for appropriate power corrections)
- Be no more different than ½ of the pooled standard deviation

We will give the “benefit of the doubt” on missing measurement information (e.g., if you cannot tell if the test over aligned, assume that it is not.) Even if giving the benefit of the doubt, please note what is missing.

Issue: **Qualitative Studies SRG**

Relevant to: Full Review

Protocol:

The SRG for qualitative studies asks for the following under “analysis”:

1. What approach was taken in analysis of the data? (Constant comparative, negative case/discrepant analysis, etc.)
2. Are the analysis and reporting appropriate for the design?
3. What interpretations and implications are provided?

*Constant comparative*—this method combines category coding with comparison of all units of meaning obtained through observation/discussant answers. The researcher examines each unit of meaning (either a topic or specific concept) to determine its distinctive characteristics. The researcher then groups units into categories. The process is one of continual refinement through the analysis, as initial categories are frequently changes, merged, or omitted, and new categories are developed.

*Negative case/discrepant data analysis*—the researcher looks for data that are negative or discrepant from the main body of data collected. These cases contradict the main category or pattern, or provide a different perspective on the category or pattern.

Examine the following when conducting analysis of qualitative work:

**Credibility**—How confident can you be in the researcher’s observations, interpretations, and conclusions? Are they believable (credible)? (This term is analogous to internal validity in quantitative research.) Credibility can be enhanced through:

1. *Methods of structural corroboration*—means through which multiple types of data are related to each other to support or contradict the interpretation and evaluation of a state of affairs (uses triangulation of methods/data).
2. *Consensus*—agreement among competent others that the description, interpretation, evaluation and thematic are correct (uses peer review or investigator triangulation).
3. *Referential or interpretive adequacy*—refers to accurately portraying the participant feedback through appropriate interpretation; researchers frequently work with participants for clarification on statements and to gather additional insight on statements through clarification.
4. *Theoretical adequacy or plausibility*—degree to which a theoretical explanation developed from the study fits the data (uses theory or interdisciplinary triangulation).
5. *Control of bias*—frequently, bias is controlled through reflexivity (the use of self-reflection to recognize one’s own biases) or negative case sampling (where researchers intentionally seek samples that disconfirm their expectations and explanations).

**Transferability**—degree to which the findings of the study can be applied or generalized to other contexts or other groups. (This term is analogous to external validity in quantitative research.) Transferability can be determined through descriptive adequacy, where the researcher provides accurate, detailed, and complete descriptions of the context and participants to assist the reader in determining transferability.

**Dependability**—Extent to which variation can be tracked or explained. (This term is analogous to reliability in quantitative research.)

You should also include the method of observation for data collection:

1. *Complete/Covert participant*—full member of the group under study without informing other members of the group (i.e., joining a street gang as a gang member).
2. *Participant as observer*—observer actively participates and becomes an insider in the event being observed so that he/she experiences events in the same way as the participants.
3. *Observer as participant*—researchers may interact with subjects enough to establish rapport but do not really become involved in the behaviors and activities of the group (i.e., sitting in on a class or workshop for welfare recipients).
4. *Complete observer*—Typically hidden from the group or may be in a public setting observing behavior (naturalistic observations).
5. *Collaborative partner*—equal partnership in the research process between the researcher and participants.

These definitions and descriptions draw on: Ary, D., Cheser Jacobs, L., Razavieh, A., & Sorensen, C. (2006). *Introduction to research in education: Seventh edition*. Belmont, CA: Thomson Wadsworth.