REVIEW PROTOCOL FOR THE OST PRACTICE GUIDE

Purpose of Guide: Provide guidance to district administrators, principals, teachers, and other staff who work with children and youth on how to select and organize **school-based out-of-school time programs aimed at improving academic achievement.**

School-based programs include those that are administered by a school or school district, as well as programs or services that are contracted by the school or school district. Such programs include on-site afterschool programs, summer school, supplemental tutoring, and other academically-focused activities that may take place out of the school building.

In order to find evidence to support specific recommendations (draft version below), the panel is considering extrapolating in some cases. If you are assigned an article to review that seems to fall blatantly outside the scope outlined in this document (for example, an evaluation of tutoring in small groups during the school day, or a program for high school students), please check whether a related recommendation has been checked in the Study Tracker. If so, then please continue with your review (also, there is a checkbox to indicate that a study fails screens but should be reviewed due to its relevance to the recommendations). If no recommendation is checked or this checkbox is not marked, then please contact Virginia Knechtel.

Draft recommendations from Panel (12/29/08):
1. INSTRUCTION
a. *Individualized/Adaptive:* Adapt program structure and content to individual learning needs (this recommendation includes aspects such as individual performance monitoring and either small group or one-on-one tutoring).
b. *Highly Engaged:* Provide high quality learning experiences that engage students with academic content.
c. *School Linkages:* Create strong linkages between the out-of-school program and the regular school day (includes both logistical suggestions for increasing communication and broader how-to steps as well).

2. DESIGN
a. *Staff Quality:* Hire well-trained and experienced staff, taking the specific needs of the target population into account (also includes professional development).
b. *Performance Monitoring:* Evaluate program implementation and effectiveness in raising student achievement with methods appropriate to program goals and resources.
c. *Attracting and Retaining Students:* Design program outreach to attract and retain students and monitor attendance to inform follow-up activities (includes logistical concerns and more general suggestions).
d. *Selecting Program Components:* Use evidence-based research to match program structure to desired student outcomes.

PROTOCOL:

Time Frame
Relevant to: Screening Criteria
Protocol: **1988–Present**

The WWC reviews studies published in the last 20 years. Reviewers should contact Liz if they have a study published before then to determine whether it should be reviewed (most likely not).

Age and/or Grade Range
Relevant to: Screening Criteria
Protocol: **Studies that focus on K–8 (ages 5–14) samples are eligible for review.** Eligible studies must at least overlap with this range, and results must be presented in such a way so that we can separate effects on grades K–8 from other grades if included.

Location and Language
Relevant to: Screening Criteria
Protocol: Restrict review to studies conducted in the **United States**.
The intervention does *not* have to be administered in English.

Study Design
Relevant to: Screening Criteria, Full Review
Protocol: **RCTs, QEDs, Regression Discontinuity Designs, and Single-Case Design studies that meet the screening criteria are eligible for full WWC review.**

Note: Correlational research may be cited by the Panel but will not go through the full review process. If you are assigned one of these articles, please fill out Stages 1 and 3 on the Main Table in the SRG to the best of your ability.

Intervention groups may be "bundled" interventions (i.e., the intervention may be multi-faceted). Also, comparisons of multiple interventions are eligible for review (i.e., the comparison group does not have to be "business as usual"). A thorough description of each intervention and/or comparison group is necessary. Similarly, comparisons of different levels of the intervention are also eligible.

Regression Discontinuity Designs (RDDs) and Single-Case Designs (SCDs) are eligible for review. Until their standards are disseminated, these studies will be to be sent for review by an expert in RDDs or SCDs.

Outcomes and Domains
Relevant to: Screening Criteria, Full Review
Protocol: Eligible outcomes should be related to either student academic achievement or student behavior.

Academic domain:
- **Achievement tests.** This category includes state-administered achievement tests or standardized achievement tests administered for research purposes, such as the Woodcock-Johnson Psycho-Educational Battery, the Peabody Picture Vocabulary Test, or the Comprehensive Test of Basic Skills.
- **Grades.** This category includes measures such as overall GPA and grades in any academic subject, including reading, math, or science.
- **Special education status.**
- **Grade retention.**

- **School attendance.** Days student attended school and school attendance rate.

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.*

Student behavioral domain:
- o **Classroom behavior.** This category includes teacher reports on *disciplinary* actions; teachers contacting parents about behavior problems, suspensions, missing recess or sitting in the hall, or having parents come to the school because of a problem.
- o **Alcohol/tobacco/other drug use.** Includes self reports.
- o **Violent/aggressive behavior.** Includes self reports.
- o **Student mental health (connected to external behaviors).** This category includes student-reported depressive symptoms, self-esteem, parent- or teacher-reported student behavior checklist, or school social behaviors scale (NOT self-esteem or "self-concepts").
- o **Homework completion.** Includes teacher or parent reports on homework completion.
- o **Level of effort.** Includes teacher reports on student effort, performance at ability level, attentiveness, participation, volunteering.

Attrition Standards for RCTs
Relevant to: Full Review
Protocol: High attrition rates (either overall or differential) can introduce selection basis. For this reason, RCTs with more than **20% overall attrition** or **5% differential attrition** (in the unit of assignment) are automatically downgraded to *meets with reservations* as their highest possible rating and are required to demonstrate baseline equivalence as discussed below.

Note: for QEDs, high attrition is less of an issue because they are already downgraded to *meets with reservations* at most and are required to demonstrate baseline equivalence. In cases of extremely high attrition for either RCTs or QEDs, studies can be failed out right. This is decided on a case-by-case basis.

Furthermore, ideally studies will perform intent-to-treat analysis. It may be difficult to distinguish if this is true, but if it is clear, be sure to note it on the Study Tracker ("intent to treat analysis" checkbox).

Baseline Characteristics for Equivalence within QEDs (and RCTs with high attrition)
Relevant to: Full Review
Protocol: Authors should demonstrate baseline equivalence on measures related to the outcome variables of interest. That is, to measure the effects of a intervention on student achievement, the intervention group and comparison group (post-attrition) should have equivalent **baseline achievement levels.** Roughly speaking, groups are equivalent at baseline if their means are within ½ of the (pooled) standard deviation of each other.

If they are not equivalent, or they did not collect the necessary data, then the study fails. If it unclear whether it was collected or if it is just not reported, then the study is rated as *uncertain*.

Ideally, groups are also equivalent in terms of **demographics** variables. If not, please be sure to note this on the SRG and on the Study Tracker, but do not downgrade the study for this.

Furthermore, QEDs and RCTs with high attrition that are capable of demonstrating baseline equivalence should control for these variables when calculating effect sizes.

Subgroup Analysis
Relevant to: Full Review
Protocol: If available, include subgroup analysis on:
- Gender
- Low vs. high achievement levels at baseline
- Special ed vs. not
- Elementary vs. Middle

Study Ratings
Relevant to: Full Review
Protocol: ***Meets standards, Meets standards with reservations, Does not meet standards,*** **and** ***Uncertain*** (*uncertain* is for cases where an author query would be necessary to determine rating—see Missing Information box below)

Missing Information
Relevant to: Full Review
Protocol: Because of the tight timeline for practice guides, generally authors are not contacted. If the **study is missing information** *necessary to determine the evidence 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 there is a chance we'll contact the author.)

We will give the "benefit of the doubt" on attrition or missing information about outcome measures. Even if giving the benefit of the doubt, please note what is missing, but do not rate as *uncertain*. We will NOT give the benefit of the doubt of baseline equivalence for achievement in QEDs or RCTs with high attrition; rate as *uncertain* (unless it *does not meet standards* for other reasons).

Description of Intervention (stage 3 on SRG)
Relevant to: Screening Criteria, Full Review
Protocol: Detailed descriptions of interventions. When possible, please note any of the following:
Intervention setting
Duration (e.g., number of minutes per session)
Frequency offered (twice a week, everyday)
Months out of the year offered
Average attendance
Attendance requirements
Who is eligible for program (anyone, low SES, low academics, etc.)
Parental involvement or requirements in program
Person who delivers the intervention (e.g., certified teacher, paraprofessional, trained former teacher working on an hourly basis)
Student to staff ratio (how is this calculated?)
Group size (i.e., one-on-one, small group, large group)
Group composition (e.g., homogenous grouping by ability level, heterogeneous groupings)

Materials/equipment used (if applicable)
Transportation provided (if applicable)
Services/activities provided (such as tutoring, instruction, homework help, computer work, snack, physical activity, mentoring, career focused, self-esteem building, etc.)
Other