

## ***About the Best Evidence Encyclopedia***

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education. It is intended to give educators and researchers fair and useful information about the strength of the evidence supporting a variety of programs available for students in grades K-12.

The Best Evidence Encyclopedia provides summaries of scientific reviews produced by many authors and organizations, as well as links to the full texts of each review. The summaries are written by CDDRE staff members and sent to review authors for confirmation.


### ***Criteria for Inclusion of Reviews***

The reviews selected for inclusion in the Best Evidence Encyclopedia are meta-analyses or other quantitative syntheses that apply consistent, scientific standards to bodies of evidence that both meet high standards of methodological quality and evaluate realistic implementations of programs currently available to educators. Specifically, to be included, reviews must:


1. Consider all studies in their area, and carry out an exhaustive search for all studies that meet well-justified standards of methodological quality and relevance to the issue being reviewed.
2. Present quantitative summaries of evidence on the effectiveness of programs or practices used with children in grades K-12, focusing on achievement outcomes.
3. Focus on studies comparing programs to control groups, with random assignment to conditions or matching on pretests or other variables that indicate that experimental and control groups were equivalent before the treatments began.
4. Summarize program outcomes in terms of effect sizes (experimental-control differences divided by the standard deviation) as well as statistical significance.
5. Focus on studies that took place over periods of at least 12 weeks, to avoid brief, artificial laboratory studies.
6. Focus on studies that used measures that assessed the content studied by control as well as experimental students, to avoid studies that used measures inherent to the experimental treatment.


### ***Basis for Program Ratings***

Educational programs in the reviews were rated according to the overall strength of the evidence supporting their effects on student achievement. "Effect size" (ES) is the proportion of a standard deviation by which a treatment group exceeds a control group. Large studies are those involving a total of at least 10 classes or 250 students. The reviews produced by CDDRE staff use the following categories of ratings:

 **Strong Evidence of Effectiveness:** At least one large randomized or randomized quasi-experimental study and one additional large qualifying study, or multiple smaller studies, with a combined sample size of 500 and an overall weighted mean effect size of at least +0.20.

 **Moderate Evidence of Effectiveness:** Two large matched studies, or multiple smaller studies with a collective sample size of 500 students, with a weighted mean effect size of at least +0.20.

 **Limited Evidence of Effectiveness: Strong Evidence of Modest Effects:** Studies meet the criteria for “Moderate Evidence of Effectiveness” except that the weighted mean effect size is +0.10 to +0.19.

 **Limited Evidence of Effectiveness: Weak Evidence with Notable Effect:** A weighted mean effect size of at least +0.20 based on one or more qualifying studies insufficient in number or sample size to meet the criteria for “Moderate Evidence of Effectiveness”.

**N No Qualifying studies:** No studies met inclusion standards.

Reviews conducted by other organizations use similar categories.

### *Interpreting Effect Sizes*

Throughout the Best Evidence Encyclopedia, the term “effect size” (ES) is used. This is the difference between the mean of the experimental group and the mean of the control group (after adjustment for any pretest differences), divided by the standard deviation of the control group. When means or standard deviations are not reported, ES is often estimated from other information that is available.

What is considered a large effect size? There is no universally accepted definition. More is better, but often the quality of the research design is more important than the size of the effect. For example, a large experiment with random assignment to treatments that obtained an effect size of +0.20 is more important than a small, matched experiment with an effect size of +0.90. Small and matched studies are more likely to have unreliable, possibly biased findings, while you can rely on the positive effect size in the large, randomized study.

One way to interpret the size of difference indicated by an effect size is to consider the improvement in percentile scores that would take place if a program with a given effect size is implemented. The table below shows this:

An effect size of... would increase percentile scores from:

An effect size of...	Would increase percentile scores from:
+0.10	50 to 54
+0.20	50 to 58

+0.30	50 to 62
+0.40	50 to 66
+0.50	50 to 69
+0.60	50 to 73
+0.70	50 to 76
+0.80	50 to 79
+0.90	50 to 82
+1.00	50 to 84

### ***Contact the BEE***

To share any comments or questions, to nominate reviews for inclusion, or for any other purpose, please contact us at:

### ***Funding for the BEE***

The Best Evidence Encyclopedia is supported by the Institute of Education Sciences, U.S. Department of Education (Grant No. R305A040082). However, any opinions expressed are those of the authors and do not necessarily represent IES positions or policies.

### ***Resources / General***

#### ***Standards for Research-Proven Programs***

Was the program compared to a control group?

- Yes
- No

Was the study at least 12 weeks in duration?

- Yes
- No

Were the program and control groups equivalent in achievement at the pretest? (within half a standard deviation)

- Yes
- No

Was the post-test a valid test, such as a standardized or state accountability test, not inherent to the program tested?

- Yes
- No

Were positive effects found in at least 2 studies in which the sample size was at least 5 classes or 125 students per treatment group?

- Yes
- No

Averaging across all studies, was the average advantage of the program group over the control group at least 20% of a standard deviation?

- Yes
- No

**The Best Evidence Encyclopedia**  
**Center for Data-Driven Reform in Education**  
**Johns Hopkins University**  
**200 W. Towsontown Blvd.**  
**Baltimore, MD 21204**  
[thebee@bestevidence.org](mailto:thebee@bestevidence.org)

### **Related References**

Slavin, Robert. "Evidence-Based Reform in Education: What Will it Take?" *European Educational Research Journal* 7.1 (2008): 124-128.  
[www.worldwords.co.uk/rss/abstract.asp?j=eerj&aid=3213](http://www.worldwords.co.uk/rss/abstract.asp?j=eerj&aid=3213)

Slavin, Robert. "Perspectives on Evidence-Based Research in Education – What Works? Issues in Synthesizing Educational Program Evaluations." *Educational Researcher* 37.1 (2008): 5-14.  
[edr.sagepub.com/cgi/content/short/37/1/5](http://edr.sagepub.com/cgi/content/short/37/1/5)