





Best Evidence Encyclopedia Program Ratings Basis


Programs are rated according to the overall strength of the evidence in support of their effects on student achievement. *Effect size* (ES) is the proportion of a standard deviation by which a treatment exceeds a control group. Large studies are those involving a total of at least ten classes or 250 students. Average effect sizes were weighted by sample sizes in computing means. The categories are as follows:

 **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 that are insufficient in number or sample size to meet the criteria for moderate evidence of effectiveness

 **Insufficient Evidence of Effectiveness:** One or more qualifying studies does not meet the criteria for limited evidence of effectiveness.

N=No Qualifying Studies: No studies met inclusion standards.

