What Works Clearinghouse

ies INSTITUTE OF EDUCATION SCIENCES

May 2012

WWC Review of the Report "Classroom Assessment for Student Learning: The Impact on Elementary School Mathematics in the Central Region"[1,2]

The findings from this review do not reflect the full body of research evidence on *Classroom Assessment for Student Learning (CASL)*.

What is this study about?

This study used a random assignment design to investigate the impact of *Classroom Assessment for Student Learning (CASL)* on elementary students' mathematics achievement.

A total of 67 schools across 32 Colorado school districts were randomly assigned to either an intervention condition that used *CASL* or a comparison condition that did not use *CASL*. The study analyzed data from 2,860 students in 33 schools with *CASL* and 3,379 students in 34 comparison schools without *CASL*. Fourth- and fifth-grade teachers in the intervention schools studied the *CASL* materials and applied *CASL* principles, practices, and tools in their classrooms during the training year. The intervention teachers then implemented the *CASL* program in their classrooms for one full school year. Teachers in the comparison group took part in their regular professional development activities.

The study assessed the effectiveness of the *CASL* program by comparing mathematics achievement of students in the *CASL* and comparison groups in the spring of the implementation year.

Features of *Classroom Assessment for Student Learning (CASL)*

CASL is a professional development program on classroom and formative assessment published by the Assessment Training Institute of Pearson Education. The *CASL* program includes a textbook, DVDs, ancillary books, and an implementation handbook, all of which are used to train teachers to conduct classroom assessments that are appropriate for, and aligned with, their learning targets.

CASL is typically implemented via teacher learning teams, in which teachers meet regularly to discuss and reflect on the content of the textbooks and DVDs and to share their experiences applying the program in their classrooms. Part of *CASL's* approach is to increase student involvement in all aspects of assessment.

This study hypothesized that use of *CASL* would increase teachers' knowledge and quality of classroom assessment practices, which in turn would lead to improved student motivation and math achievement.

WWC Rating

The research described in this report meets WWC evidence standards without reservations

Strengths: This study is a well-implemented randomized controlled trial.

What did the study find?

The study found no effects of *CASL* on the mathematics achievement of fourth- and fifth-grade students. The estimated effect size of 0.01 is neither statistically significant nor substantively important.

Appendix A: Study details

Randel, B., Beesley, A. D., Apthorp, H., Clark, T. F., Wang, X., Cicchinelli, L. F., & Williams, J. M. (2011). *Classroom Assessment for Student Learning: The impact on elementary school mathematics in the Central Region* (NCEE 2011-4005). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Setting

The study was conducted in fourth- and fifth-grade classrooms in 67 public elementary schools spread across 32 districts in Colorado.

Study sample

Public schools with at least one teacher in fourth grade and one teacher in fifth grade were recruited for the study. Of the estimated 981 eligible schools in the state, 67 schools (7%) volunteered to participate. Schools were blocked by district or "pseudo-district" and randomly assigned to the intervention or comparison condition. Schools in six districts were randomly assigned within their district. When there was only one study school in a district, schools were included in a "pseudo-district" block of similar schools (in terms of location and date of entry into the study), and random assignment was done within this block. If a block contained an odd number of schools, the odd school was assigned to the comparison group, resulting in 33 schools in the intervention group and 34 schools in the comparison group. One intervention school and two comparison schools dropped out of the study during the orientation year but were included in the analyses. On average, the student population of the sample schools included 44% minority students and 47% students eligible for free or reduced-price lunch.

The study sample included students who were in fourth- or fifth-grade during the implementation year (2008–09). At the time of random assignment, 4,420 students were in intervention classrooms, and 5,176 students were in comparison classrooms. Student absences and mobility over the course of the study resulted in missing pretest and/or posttest data from approximately 35% of the students; the resulting analysis sample included 2,860 intervention students and 3,379 comparison students.

Intervention group

Intervention teachers studied the *CASL* materials and applied *CASL* principles, practices, and tools in their classrooms during the training year (2007–08). After completing the training year, intervention teachers then implemented the *CASL* program in their classrooms for one full school year (2008–09).

Comparison group

Teachers in the comparison group took part in their regular professional development activities. Comparison schools were provided with financial resources approximately equivalent to the cost of the *CASL* materials but were not provided with specific materials.

Outcomes and measurement

The student outcome measure was the mathematics subtest of the Colorado Student Assessment Program (CSAP) standardized assessment. Spring 2007 pretest scores were used for fifth-grade students, and spring 2008 pretest scores were used for fourth-grade students. These pretest scores were included as covariates to adjust for pretest differences in the analytical model. The posttest was administered to both cohorts in spring 2009, and test scores for both cohorts of students were combined. For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

Teachers in the intervention group received a complete set of *CASL* professional development materials in fall 2007, including a facilitation handbook, *CASL* textbooks, DVD sets, and ancillary books. Formal training involved an introductory videoconference and access to district staff who were trained as facilitators. Teachers were asked to form learning teams to study the *CASL* materials during the 2007–08 school year and to fully implement the *CASL* practices in 2008–09. Implementation findings indicate that learning teams were formed in most schools; approximately 63% of teachers attended the recommended nine learning team meetings called for by *CASL*. The average amount of time teachers reported spending on *CASL* training was 31 hours, compared with the 60 hours recommended by the program's developer.

Identification

This study is an Institute of Education Sciences (IES)-funded study that was conducted by Regional Educational Laboratory (REL) Central.

Appendix B: Outcome measure for the mathematics achievement domain

Mathematics achievement	
Colorado Student Assessment Program (CSAP) mathematics subtest	The CSAP is the statewide achievement test used to measure adequate yearly progress under the No Child Left Behind (NCLB) Act. The test is aligned with the Colorado Model Content Standards and Assessment Framework and is vertically scaled from grades 3 through 8. Internal consistency on the mathematics test for grades 4 and 5 is 0.94, with alpha coefficients ranging from 0.84 to 0.95 for NCLB subgroups in those grades.

Appendix C: Study findings for the mathematics achievement domain

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Mathematics achievement								
CSAP mathematics subtest	Grades 4 and 5	67 schools/ 6,239 students	514.66 (76.40)	513.02 (80.22)	1.64	0.01	0	0.91
Domain average for mathematics achievement						**0.01**	**0**	**Not statistically significant**

Table Notes: Positive results for mean difference, effect size, and improvement index favor the intervention group; negative results favor the comparison group. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the change (measured in standard deviations) in an average student's outcome that can be expected if the student is given the intervention. The improvement index is an alternate presentation of the effect size, reflecting the change in an average student's percentile rank that can be expected if the student is given the intervention. The statistical significance of the study's domain average was determined by the WWC; a study is characterized as not statistically significant when univariate statistical tests are reported for each outcome measure, and each of the effects within the domain are not statistically significant. CSAP = Colorado Student Assessment Program.

Study Notes: One intervention school and two comparison schools dropped out of the study during the orientation year; data from these schools were included in the results of the analytical model reported here. The study estimated the effect of *CASL* on student mathematics achievement with a two-level model in which students were nested within schools. The main model presented in the study used the expectation maximization algorithm with multiple imputation to impute missing pretest and posttest data from approximately 35% of the students. Models were also estimated using casewise deletion for missing values. No significant results were found using any analytic technique or way of handling missing data. The student sample sizes, means, standard deviations, effect size, and *p*-value are based on nonimputed data provided to the WWC by the authors. The means and standard deviations are not adjusted for clustering. The WWC calculated the intervention group mean using a difference-in-differences approach (see the WWC Procedures and Standards Handbook, Appendix B) by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. The effect size, improvement index, and *p*-value come from a two-level mixed model that controls for pretest scores and for clustering of students within schools. No corrections for clustering or multiple comparisons were needed.

Endnotes

[1] Single study reviews examine evidence published in a study (supplemented, if necessary, by information obtained directly from the author[s]) to assess whether the study design meets WWC evidence standards. The review reports the WWC's assessment of whether the study meets WWC evidence standards and summarizes the study findings following WWC conventions for reporting evidence on effectiveness. The WWC rating applies only to the summarized results, and not necessarily to all results presented in the study. This study was reviewed using the Elementary School Mathematics review protocol, version 2.0.

[2] Absence of conflict of interest: The Regional Educational Labs were provided technical assistance by Mathematica Policy Research, which also operates the WWC. For this reason, this study was reviewed by staff from subcontractor organizations.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2012, May). *WWC review of the report: Classroom Assessment for Student Learning: The impact on elementary school mathematics in the Central Region*. Retrieved from http://whatworks.ed.gov.

Glossary of Terms

Attrition Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.

Clustering adjustment If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.

Confounding factor A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.

Design The design of a study is the method by which intervention and comparison groups were assigned.

Domain A domain is a group of closely related outcomes.

Effect size The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.

Eligibility A study is eligible for review and inclusion in this report if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.

Equivalence A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol.

Improvement index Along a percentile distribution of students, the improvement index represents the gain or loss of the average student due to the intervention. As the average student starts at the 50th percentile, the measure ranges from –50 to +50.

Multiple comparison adjustment When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.

Quasi-experimental design (QED) A quasi-experimental design (QED) is a research design in which subjects are assigned to intervention and comparison groups through a process that is not random.

Randomized controlled trial (RCT) A randomized controlled trial (RCT) is an experiment in which investigators randomly assign eligible participants into intervention and comparison groups.

Single-case design (SCD) A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.

Standard deviation The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample tend to be spread out over a large range of values.

Statistical significance Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < 0.05$).

Substantively important A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Please see the WWC Procedures and Standards Handbook (version 2.1) for additional details.