Title: Student and Teacher Impacts of Professional Development in Classroom Assessment for Student Learning

Author(s):

Bruce Randel
Andrea D. Beesley
Tedra F. Clark
Xin Wang
Abstract Body

Background

Under the No Child Left Behind Act (NCLB), states, districts, and schools are required to ensure that all students meet the same high standards in mathematics and reading by the end of the 2013-2014 school year (No Child Left Behind Act, 2001). In working towards this goal, states, districts, and schools are increasingly in need of rigorous, high-quality research on efficient and effective interventions to improve student achievement.

This study was conducted by the Central Region Educational Laboratory (REL Central) administered by Mid-continent Research for Education and Learning (McREL) to provide educators and policymakers with rigorous evidence about the potential of a widely-used, research-based professional development program in classroom and formative assessment—Classroom Assessment FOR Student Learning (CASL)—to improve student achievement.

REL Central identified priority needs in the region for education research and technical assistance through a comprehensive review of state demographic and education system data, interviews with chief state school officers and key state agency staff, and ongoing contact with a variety of constituent groups, including policymakers, principals, and superintendents. Through the integration of these efforts, REL Central identified several priority needs to guide its work in the Central region, including a need for guidance on research-based classroom practices and a need for improved teacher quality, particularly in light of the “highly qualified” teacher requirement of the No Child Left Behind Act.

The study was designed to respond to specific regional issues and concerns expressed by chief state school officers from the Central Region states (Colorado, Kansas, Missouri, Nebraska, North Dakota, South Dakota, and Wyoming). This study addressed the regional needs of poor performance in mathematics, the lack of the use of formative assessment, and a need for quality professional development for educators. The ultimate goal of providing scientifically based guidance on formative assessment was to help schools meet Adequate Yearly Progress requirements under NCLB.

Purpose

Despite CASL’s wide usage, there is no direct causal evidence supporting its effectiveness in raising student achievement or improving other student and teacher outcomes. This study was designed to provide an unbiased estimate of the effectiveness of CASL to improve student achievement and other student and teacher outcomes. This study estimated the impact of CASL under conditions that would typically occur had the schools purchased CASL and implemented it without monitoring or involvement of research staff.

This study examines the impact of CASL on the primary outcome of student achievement in mathematics. According to the hypothesized theory of action that guided the design of this study, additional outcomes were included to address the impact of CASL on several intermediate outcomes and to provide contextual information to aid in the interpretation of the impact on the primary outcome. First, CASL was hypothesized to impact student motivation. In addition, CASL is unlikely to impact student achievement or motivation without first having an impact on teachers’ understanding and practice of formative assessment. The study, therefore, also addressed the impact of CASL on teacher knowledge of classroom assessment practices, the
quality of classroom assessment practices, and the extent to which teachers involved their students in formative assessment.

Setting

Schools were recruited from across Colorado to participate in the study. Colorado was chosen as the target state primarily because it has one of the largest populations in the Central region from which to recruit schools and because the statewide achievement test is vertically scaled. The target population for this study was public schools in Colorado large enough to have at least one Grade 4 teacher and one Grade 5 teacher. The districts and schools were located across Colorado and provided a diverse sample in terms of district and school size, levels of urbanicity (urban, suburban, town, and rural), and student achievement.

Participants

Sixty-seven schools from 32 districts in Colorado volunteered to participate and were randomly assigned to either the intervention group (CASL) or to the control group where teachers conducted their regular professional development activities. The final student impact analysis sample included 9,596 students from the study schools. Sample size was sufficient to provide statistical power (> .80) to detect an impact on student mathematics achievement of approximately .25 standard deviations. Random assignment was blocked by district and resulted in two groups of schools that were found to have no statistically significant differences on a number of characteristics, including mathematics achievement, teacher-student ratio, and percentages of students in racial/ethnic groups.

The study included the students and teachers from all the Grade 4 and 5 classrooms in the participating schools to allow for availability of baseline student achievement data from the Grade 3 administration of the statewide NCLB achievement test. More than 400 Grade 4 and 5 teachers from a variety of large and small, urban, suburban, and rural schools participated in the study. The intervention and control teachers did not differ by a statistically significant margin in terms of their education or their scores on the teacher baseline measure of assessment knowledge. Intervention teachers, however, did have more years of experience teaching and more years of experience in teaching mathematics than the control teachers; these differences were statistically significant and controlled in the impact analysis by including these two teacher experience variables as covariates.

Intervention

REL Central provided schools in the intervention group with the complete set of CASL professional development materials at the beginning of the study (November 2007), including a facilitation manual, CASL textbooks (Stiggins, Arter, Chappuis, & Chappuis, 2004) for every participating teacher, DVD sets, and ancillary books. Teachers in the intervention group also participated in an introductory videoconference with CASL author Richard Stiggins and had access to a facilitator who had attended a training workshop conducted by the CASL developers.

The CASL program is designed to be self-executing. The facilitation manual provides guidance and developer recommendations for implementing the program. The developers recommended implementing CASL via teacher learning teams, in which teachers meet regularly.
to discuss and reflect upon the content of the program provided in the textbook and DVDs and share their experiences applying the program practices and principles in their classrooms. Teachers in the intervention group implemented CASL naturally, without any involvement of, or requirements from, the research team.

Research Design

This study used a cluster randomized design in which schools were randomly assigned to either the intervention or control group to allow teacher teaming and collaboration to continue across and within grades via learning teams. Principals and other instructional leaders in schools in both the intervention and control groups were able to continue their usual practices of encouraging input and collaboration. Random assignment of whole schools also reduced the risk of cross-overs between classrooms in the same building (e.g., control-group members acquiring intervention materials and using them in their classroom). In the intervention group, teachers formed learning teams, received the professional development materials, and implemented the CASL program. Teachers in schools assigned to the control group participated in their regular professional development activities.

Data Collection and Analysis

Data were collected throughout the course of the study to describe the fidelity of CASL implementation and the larger professional development context in the study schools and to estimate the impacts of CASL on the student and teacher outcomes. Student achievement data were obtained directly from the state department of education for the CSAP administered in April of each year. Survey data were collected from teachers and students in the participating schools.

Students’ scale scores on the 2009 administration of the mathematics portion of the CSAP were used to estimate the impact of CASL on student achievement. Using the state assessment as the measure of the primary outcome—student achievement—allowed for estimation of the extent to which implementation of CASL impacted student achievement in relation to the goals of NCLB. In addition, students in participating schools were thought to be more likely to be motivated to perform well on the statewide assessment than on a separate achievement test administered solely for the purposes of this study.

Student motivation was measured with the Survey of Student Motivation, comprising of the Ongoing Engagement and Perceived Autonomy (Self-Regulation) subscales of the elementary student Research Assessment Package for Schools (RAPS-SE; IRRE, 1998) along with the Academic Efficacy subscale of the Patterns of Adapted Learning Scales (PALS; Midgely et al., 2000). The Survey of Student Motivation included 20 items and was administered in a posttest-only design in order to reduce the data collection burden on students. The RAPS-SE and the PALS were on a Likert-scale format in which students responded to numeric score categories from 1 to 4. The mean score across all 20 items was used as the measure of student academic motivation in the impact analysis.

The Test of Assessment Knowledge was developed for this study and included 60 multiple-choice, true–false, and matching items. The test items covered teachers’ knowledge of, and reasoning skills regarding, generally accepted principles and practices of classroom assessment. Although the test was designed to be sensitive to the CASL program, steps were taken to ensure
that the test was not over-aligned; it did not include materials, text, idiosyncratic wording, or terminology from the CASL program or program materials. The test used common language and sampled from the general domain of classroom assessment.

The measure of teacher assessment practice for this study was an artifact-based instrument adapted by the study team from a work sample instrument developed at the National Center for Research on Evaluation, Standards, and Student Testing (Matsumura, Patthey-Chavez, Valdes, & Garnier, 2002). The Assessment Work Sample consisted of written instructions and a rubric. Packets containing the instructions and envelopes for collecting and returning artifacts were delivered to participating teachers prior to data collection. Per the written instructions, teachers collected and submitted artifacts for three types of assessment (homework/seatwork, quiz, and performance assessment); artifacts included four graded student papers and a cover sheet describing the goal of the assessment and the assessment method for each of the three types of assessment. The work samples were scored independently by two raters blind to the teachers’ experimental group membership, and the final score was the average of each rater’s score. Interrater reliability coefficients for the posttest work samples calculated as the correlation between rater scores ranged from .66 to .82.

Student involvement was measured with a teacher report. Teachers were asked to record the number of days during the previous two-week period during which they involved their students in assessment-related activities. Teachers also were asked to record the total number of instructional days for that two-week period. The survey included 14 items addressing activities such as discussing the learning objectives, evaluating their own work using scoring guides or rubrics, and revising work to correct errors. Each item from the survey was re-coded by dividing the item score (number of days for the activity) by the total number of instructional days so that the recoded score reflected the percentage of instructional days during which teachers involved their students in each respective classroom assessment activity. This was done to account for the possibility that the teachers’ two-week reporting period may have included fewer than 10 instructional days because of in-service or other non-instructional time. The total score on the survey was the mean percentage across all 14 items.

The primary purpose of the analysis was to provide an unbiased estimate of the impact of CASL on student achievement in mathematics. Consistent with the random assignment of schools to either the intervention or control group, the impact was estimated at the school level, using a mixed-model approach to account for the sources of variability in the data that resulted from the nested structure of the school environment.

The student achievement impact analysis sample included all schools that were randomly assigned at the beginning of the study to either the intervention or control group. The impact analysis samples for the intermediate outcomes were reduced by school and teacher attrition, but comparisons of the baseline characteristics of the sample of schools included in the intermediate outcome impact analyses did not reveal any statistically significant differences. Data were missing in each of the impact analyses samples, and the expectation maximization (EM) algorithm with multiple imputation method was used to impute missing data.

A two-level model in which students (or teachers) were nested within schools was used to estimate the impact of CASL on students’ mathematics achievement. Classroom-level teacher effects were not included in the analysis of student achievement for several reasons. First, the intervention was implemented over two years, so students were exposed to two teachers over the course of the study. Second, the student achievement data did not include information linking students to teachers or any information identifying students (other than an encrypted
identification number), so grouping students by classroom was not possible. Finally, simulation studies have found that clustering within intermediate units has little effect on Type I error (Murray, Hannan, & Baker, 1996).

Findings

In terms of implementation fidelity, the learning teams in sixty-eight percent of intervention group schools met the quality criteria for learning teams established by the CASL developers. These criteria include meeting regularly, establishing assignments between meetings, and sharing a common purpose. Ninety percent of learning teams were the recommended size of three to six members. Sixty-three percent of responding teachers attended at least the recommended nine learning team meetings. Seventy-eight percent of responding teachers reported that their learning team meetings involved discussions about what they were learning about classroom assessment. Forty-two percent of teachers reported at least partially reading each CASL textbook chapter. The average amount of total time that teachers reported spending on CASL training was 31 hours, compared to 60 hours recommended by the developer.

In terms of impacts on the primary and intermediate outcomes, impact analyses have been completed, and the final report is current under review at the IES Standards and Review Office. Approval of the report is expected in early 2011.

Conclusions

This cluster randomized trial of the CASL professional development program had sufficient statistical power to detect an impact of at least .25 standard deviations on student achievement. An intent-to-treat analysis was conducted to estimate the impact of CASL on student achievement; all schools were included in the analysis and were analyzed as randomized, regardless of the level of implementation of the intervention. Results from sensitivity analyses revealed that the impact estimates on student achievement were robust to decisions regarding the inclusion of covariates, estimation method, and the treatment of missing data. In other words, design and analysis decisions made by the research team did not change whether or not the impact results would have been statistically significant.

Interpretation of this study is subject to several limitations. First, results generalize only to implementation of CASL at levels similar to those found in this study, to the voluntary sample included in the study, and to students’ Grade 4 and 5 mathematics achievement as measured by the Colorado statewide achievement test. Second, although attrition was not an issue for the student achievement outcome, the non-response for the student motivation outcome and the teacher outcomes, however, exceeded levels considered acceptable by the What Works Clearinghouse (What Works Clearinghouse, 2009). The multiple imputation method used to impute missing data for the teacher outcomes resulted in a teacher impact analysis with levels of attrition expected to result in an acceptable level of bias (What Works Clearinghouse, 2009).
Appendices

Appendix A. References


Appendix B. Tables and Figures

NA