Abstract Title Page
Not included in page count.

Title: The Impact of the Measures of Academic Progress (MAP) Program on Student Reading Achievement

Authors and Affiliations:
David S. Cordray, Vanderbilt University
Georgine M. Pion, Vanderbilt University
Chris Brandt, REL Midwest at the American Institutes for Research
Ayrin Molefe, REL Midwest at the American Institutes for Research

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Effective differentiation of instruction based on student readiness and learning profiles requires valid descriptive data at the classroom level (Decker 2003). Although teachers may use their own student-level assessments (tests, quizzes, homework) to monitor learning, it is challenging to use performance on classroom measures to assess likely performance on external measures, such as statewide tests or nationally normed standardized tests. Benchmark measures reflective of such external tests may be more useful in helping teachers make decisions about differentiating instruction, which in turn can lead to gains in student learning and higher scores on state standardized tests (Baenen, et al. 2006; Baker and Linn 2003).

One of the most widely used commercially available systems incorporating benchmark assessment and training in differentiated instruction is the Northwest Evaluation Association's (NWEA) Measures of Academic Progress (MAP) program. The MAP program involves two components: 1) computer-adaptive assessments administered to students three to four times per year, and 2) teacher training and access to MAP resources on how to use data from these assessments to differentiate instruction. The MAP program is currently used in over 20% of school districts nationwide (http://www.nwea.org/support/article/1339).

Although the MAP program is used extensively in school districts across the United States, there is no experimental evidence of its impact on student outcomes. Given that the number of schools investing in MAP and similar programs is projected to increase, rigorous evidence of the effectiveness of such programs is critical. This goal of this study is to conduct a rigorous evaluation of the impact of the MAP program on teachers' differentiated instructional practices and students' reading achievement in grades 4 and 5.

Purpose / Objective / Research Question / Focus of Study:

This report focuses on the program's impact after the second year of implementation,[1] and seeks to answer the following questions on implementation fidelity and student outcomes:

1. Were MAP resources (training, consultation, web-based materials) delivered by NWEA and received and used by teachers as planned?
2. Did MAP teachers apply differentiated instructional practices in their classes to a greater extent than their control counterparts?
3. Did the MAP program (that is, training plus benchmark testing feedback) affect the reading achievement of grades 4 and 5 students after the second year of implementation, as measured by the Illinois Standards Achievement Test (ISAT) reading scale scores or the MAP composite test scores in reading and language use?
4. Were there variations in the impacts of the MAP intervention on grades 4 or 5 ISAT reading and MAP composite scores across subgroups of students after the second year of implementation?

Setting:

The study focused on grade 4 and 5 students in 32 public elementary schools across five districts in Illinois. To be eligible, schools needed to have at least one full-time regular classroom

[1] The MAP training consists of four one-day sessions throughout the school year. The most critical of these sessions—Session 3 on using MAP to differentiate instruction—was not delivered until January 2009 (in the first

teacher who taught reading in a self-contained classroom in grade 4 and one full-time regular classroom teacher who taught reading in a self-contained classroom in grade 5.

Population / Participants / Subjects:

Of the 32 schools enrolled in the study, 28 (87.5 percent) were eligible for Title 1 services, and 78.1 percent were located in either a city or a suburb.[2] On average, about half the students in the participating schools were eligible for free or reduced-price lunch (range: 0–95 percent), and about 62 percent were White (range: 8–97 percent). Total enrollment in the study schools ranged from 162 to 701, with an average of 385 students (including about 60 students in grade 4 and 60 in grade 5) taught by about 23 full-time teachers in each school. A total of 172 teachers (85 grade 4 teachers, 87 grade 5 teachers) and 3,720 (1914 grade 4 students; 1806 grade 5 students) students were in the final analytic sample.

Intervention / Program / Practice:

The MAP program has two main components: an extensive portfolio of tests and training and on-demand support in the use of test results to guide instructional practice. The MAP assessments are a collection of computer-adaptive tests that place individual students on a continuum of learning from grade 3 to grade 10.[3] Schools and teachers typically administer the test three times per school year and use MAP results to monitor their students' progress toward state proficiency standards. Because the tests are computer-adaptive, students are given their overall score immediately after the test ends, and teachers can generate a series of customized reports within 24 hours of administration. MAP training consists of four one-day training sessions, along with on-demand consultation through conference calls and on-site visits from an NWEA MAP coach throughout the school year. The primary objectives of the training are to equip teachers with the knowledge and skills to administer the tests; generate and interpret outcome reports at the individual, group, and classroom level; use report results and other MAP online resources to determine student readiness and differentiate instruction; and use MAP data over time to set student growth goals and evaluate instructional programs and practices.

Research Design:

This study used a cluster-randomized design that randomly assigned 32 schools from five Illinois districts to implement the MAP program at either fourth or fifth grade. If grade 5 classrooms in School A were assigned to the treatment condition, grade 4 classrooms in the school were assigned to the control condition. If grade 5 classrooms in School B were assigned to the control condition, grade 4 classrooms were assigned to the treatment condition. The control group for grade 4 classes consisted of grade 4 classes in schools in which MAP was randomly assigned to grade 5, and the control group for grade 5 classes consisted of grade 5 classes in schools in which MAP was randomly assigned to grade 4. This randomization technique resulted in two experiments, one at grade 4 and one at grade 5, and produced a valid counterfactual for the treatment group within each grade (see Borman et al. 2007 for a similar

---

[2] These classifications are based on the National Center for Educations Statistics revised (2006) typology of locale codes, in which city, suburb, town, and rural were subclassified into three categories, resulting in 12 urban locale codes (http://nces.ed.gov/ccd/rural_locales.asp).

[3] For this study, the researchers employed the MAP tests in reading and language usage for grades 4 and 5 and administered the tests three times a year (in the fall, winter, and spring) to treatment students and once (in the spring) to control students.

randomization design).[4] Randomization of schools was stratified by district, and analyses were conducted separately for grade 4 and grade 5, and separately for the two outcomes: ISAT reading and MAP composite scores.

Data Collection and Analysis:

Multiple data collection methods were used to describe and assess MAP implementation fidelity. MAP administrative records (such as training attendance data) and web-based computerized reports were used to describe the extent to which NWEA delivered the program to the study schools as intended. Teacher surveys, instructional logs, and classroom observations were used to assess whether teachers in the treatment group implemented core components underlying the MAP training (for example, differentiated instruction practices) to a greater extent than their control group counterparts. Students' reading performance was assessed with the spring 2010 Illinois Standards Achievement Test (ISAT) in reading. The ISAT is administered every spring to all Illinois students in grades 3–8. In addition, results of the spring 2010 MAP tests in reading and language usage were used as a composite measure to assess students' reading and literacy achievement.

Intent-to-treat estimates of impacts on student outcomes were obtained using two-level hierarchical regression models to adjust for the clustering of students within schools, and district fixed effects to control for the randomization of schools within districts. The models also incorporated baseline student characteristics (prior reading achievement, gender, socioeconomic status, racial/ethnic minority status, English proficiency status, and disability status); teacher characteristics (gender, graduate degree status, teaching experience in English language arts, licensure status, racial/ethnic minority status); and school mean prior reading achievement on the ISAT. The overall impacts are presented as averages of district-specific impacts obtained from the regression models, weighted by the number of schools in each district. The analytic sample consisted of 1,914 eligible grade 4 students (and 85 grade 4 teachers) and 1,806 grade 5 students (and 87 grade 5 teachers) from 32 participating schools using multiple imputation to fill in missing outcome and covariate values.[5]

Findings / Results:

Implementation by NWEA and MAP Teachers: NWEA provided the resources needed to support the MAP program at the school and classroom levels. Throughout the study period, testing resources were fully available in all schools, web-based resources were continuously available, and MAP training and testing were scheduled and conducted in a timely fashion. During both years of the intervention NWEA trainers were available for follow-up consultations. Implementation of the MAP program unfolded without any notable problems.

The study team identified 12 MAP–relevant components that teachers could implement during the two-year period of this study. The same implementation profile was observed for grade 4 and grade 5 MAP teachers. Participation rates varied across the 12 program components, ranging from 36 percent (use of MAP web-based resources) to 90 percent (use of MAP resources for planning lessons). There was considerable variation in the dose level across teachers (ranging from 0 to 100%). The average dose of MAP program components was 66% in both grades.

[4] The counterfactual condition included schools that implemented a variety of assessment types but had never implemented benchmark assessment or conducted training to help teachers interpret and use benchmark data to inform their instruction.

[5] This sample includes one school that dropped out of the study immediately after randomization.

About half of teachers participated at rates of 75% or higher. The dose data suggest that there was substantial variability in the extent to which MAP teachers implemented the program.

Use of Differentiated Instruction: Data from classroom observations and teacher logs show small, nonsignificant differences in the use of key aspects of differentiated instruction. However, teacher reports of differentiation in grade 5 reveal differences between conditions. The grade 5 differences were statistically significant for the survey composite measure ($p < .001$) and the achieved relative strength index (ARSI)[6] was relatively large (0.894). The survey composite for grade 4 was not significant at $p < .05$, and the ARSI was modest (0.335). The best estimate of the ARSI for differences between conditions across the three measures was 0.227 for grade 4 and 0.340 for grade 5. By conventional standards for interpreting effect sizes, these estimates reflect small differences.

Student Achievement: The MAP program had no statistically significant overall impact on the reading achievement of grades 4 or 5 students as measured by the ISAT reading scale score or the MAP composite scores. The directions (positive) and magnitudes of the impacts were similar for the two outcomes: a 0.05 standard deviation for the ISAT reading score and a 0.07 standard deviation for the composite MAP score. Statistically significant subgroup and differential impacts were observed at grade 4 and grade 5. Specifically, at grade 4, the intervention had a positive impact on the ISAT reading scores of high socioeconomic status students and students' with high initial reading ability. The intervention had a differential impact on both the ISAT reading and MAP composite scores of students whose initial reading ability was low and students whose initial reading ability was high. At grade 5, there were no statistically significant impacts on the subgroups examined; however, there was a differential impact between White and racial/ethnic minority students on the MAP composite score.

Conclusions:

The increasing demand for the MAP program in particular, and benchmark assessments in general, not only in the Midwest but across the nation warrants the need for a rigorous evaluation of the MAP program. This study not only adds to the existing literature on professional development and benchmark assessments but more importantly, provides rigorous evidence on the effectiveness of programs that incorporate both.

The differential impact among low and high ability students at grade 4 suggests that the MAP program may have the greatest impact on low and high ability students. More research is needed to confirm a causal relationship between the MAP Program and increased performance among high and low ability students. In addition, research should examine how the assessments might be interpreted and used by teachers in qualitatively different ways, or perhaps to a greater extent, to address these students' learning needs.

[6] ARSI is the difference in group averages on the fidelity index divided by the pooled standard deviation for that index (Cordray and Pion (2006) and Hulleman and Cordray (2009)).

Appendices

Not included in page count.

Appendix A. References

Baenen, N., Ives, S., Lynn, A., Warren, T., Gilewicz, E., & Yaman, K. (2006). *Effective practices for at-risk elementary and middle school students* (E&R No. 06.03). Raleigh, NC: Wake County Public School System.

Baker, E. L., & Linn, R. L. (2003). Validity issues for accountability systems. In S. H. Fuhrman & R. F. Elmore (Eds.), *Redesigning accountability systems for education* (pp. 47–72). New York: Teachers College Press.

Borman, G. D., Slavin, R. E., Cheung, A. C. K., Chamberlain, A. M., Madden, N. A., & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of Success for *All*. *American Educational Research Journal*, *44*(3), 701–731.

Cordray, D. S., & Pion, G. M. (2006). Treatment strength and integrity: Models and methods. In R. Bootzin & P. McKnight (Eds.), *Strengthening research methodology: Psychological measurement and evaluation*, 103–124. Washington, DC: American Psychological Association.

Decker, G. (2003). Using data to drive student achievement in the classroom and on high-stakes tests. *T.H.E. Journal, 30*(6). Retrieved December 14, 2010, from http://www.thejournal.com/articles/16259

Hulleman, C., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Educational Effectiveness, 2*(1), 88–110.

Appendix B. Tables and Figures

Not included in page count.