Making the Most of Available Data: Descriptive Trends in Student Impact from the Missouri Improving Teacher Quality Program

Martha A. Henry, Ed.D., C.E.P., & Keith S. Murray, C.E.P.,
MA Henry Consulting, LLC
(For information, contact mahenry@mahenryconsulting.com, 703-204-4760)

Abstract

Understanding the impact of teacher professional development programs such as the federally funded Missouri Improving Teacher Quality Program Grant Program requires both appropriate data and appropriate analytic scrutiny among evaluators charged with the task. After having applied rigorous propensity-score matching to investigate statistically significant differences in standardized test scores between students grades 3-12 (N = 19,539 student math and science tests) of program teachers and a control group, the authors returned to the available data to investigate trends using descriptive methods. Analysis of trends for teachers with at least two years of participation by student grade/test results indicated generally positive indications of enhanced achievement for nearly all grades. Despite the inability to conclude causal linkages between the program's intervention with teachers and student results, correlative indications are important in expanding the context of evaluation of program effects. Making the most of available data through responsible and robust investigations – distinct from haphazard "data mining" adds value to evaluation and understanding of program impact. The need to sustain data availability and ensure the quality of data from standardized test sources is crucial, as is the continuation of such teacher professional development programs through best-practice, competitive funding, as has been possible through the Improving Teacher Quality Grant Program.

Introduction

For a program such as the U.S. Department of Education Title II, Part A *Improving Teacher Quality Grant Program (ITQG)*, the stakes for connecting targeted interventions on mathematics and science teacher performance to subsequent student outcomes are understandably high and unsurprisingly challenging. The program, implemented on a formula basis in each state via state education, higher-education and local education agencies through competitively-funded university-based projects, reaches all grade levels

from kindergarten through 12th grade, with a special emphasis on high-need school districts. The funding represents the largest federal initiative for teacher professional development projects (Federal Education Budget Project, 2015). In Missouri's ITQG, the focus of this paper, the program intends to enhance teacher content knowledge, pedagogical practice and assessment skills, and thereby improve student achievement in the subjects addressed. It has been administered by the Missouri Department of Higher Education through 12 annual cycles to date, with at least one more cycle in preparation.

Evaluating the impact of such a teacher professional development program on student achievement requires a complex and rigorously designed data collection and analysis plan. Establishing or even suggesting causal links is especially daunting with such evaluations (Aldeman et al., 2011). A broad range of grade levels and content foci, as well as geographic diversity, individual projects' implementations, and the practical and logistical limitations also are involved, additional to causal-analysis challenges (Miles, et al., 2004).

Apart from potential influences affecting both teacher and student performance other than the ITQG program, applying appropriate measures to student achievement is difficult. Individual project-developed student tests often lack sufficient validation and reliability testing, which can require time and resources beyond those available. Student tests also must address the variety of grades included in each project. Project-derived tests may be so specific to project-based content areas that a narrow interpretation of achievement enhancement may result (Blank & de la Alas, 2009, pp 16-18).

Attempting to apply extant national tests or test items as an alternative, even when they are well validated and reliable, can come at the cost of alignment to specific project content. Confusion of standard student assessment practice with evaluative needs,

comprising a different discipline, can occur, especially among non-evaluators charged with evaluative responsibilities. Perhaps most challenging is the difficulty of generalizing the results of evaluation of individual projects with the need for external evaluation to take a programmatic view across all projects in a cycle. Issues of sample size, statistical power and other concerns add to the considerations.

An alternative has been developed and tested over two cycles by the authors of this paper. This plan has involved applying statewide standardized grade-level test data (mathematics, grades 3-8; science, grades 5 and 8) and end-of-course Biology I and Algebra I regardless of grade level) from the Missouri Assessment Program (MAP), available from the Missouri Department of Elementary and Secondary Education. Analyses of test scores for students of participating teachers and a statistical control group resulting from propensity-score matching offered a basis for considering the effects of the ITQG program on students in a rigorous manner (Henry, Murray & Hogrebe, 2013). While these data sets are developed in each state, their application for such purposes appears to have remained infrequent.

Even so, issues of sample size by grade and the lack of data for students in the earliest grade levels do affect the completeness of results. Incorporating multiple years of data for trend analysis and the uncertain quality-control measures in place also has complicated the process. However, such an approach continues to offer the most complete and objective measures for evaluatively considering the effect of the ITQG program on students, given the complexity of the situation. Standardized-test data by their nature express the state's official expectations of student achievement in appropriate subjects.

The results of these analyses have been reported in various documents for cycles 9 and 10 of the ITQG. (Henry, Murray & Hogrebe, 2012; Henry, Murray & Hogrebe, 2013). The analyses considered student test scores for the years 2009-2012, and also are briefly described in this paper. These data were not available for Cycle 11 evaluation, and an adoption of extant, publicly available data by school district and grade level was applied in reporting for that year (Henry, Murray & Hogrebe, 2014).

This introduction has been necessary to provide the context for the present paper, which is to further consider applications of standardized test data for evaluation purposes. Given the need for access to data, stripped of identifiers but individually coded and processed in order to fulfill the requirements of comparison-group matching and other analytic details, the authors have wanted to see if other analyses of the data currently available to them could add to the understanding of the effect of ITQG teacher treatment on students. Continuing the investigation of the possible incorporation of MAP's standardized-test data for program evaluation, the authors have descriptively considered trends among teachers participating in an ITQG project over at least two years.

Methods

Previous analysis of MAP data from the years 2009-2012 (pre-treatment 2009-2010, treatment 2011-2012) showed statistically significant gains ($p < .05$) for Cycle 9 (2011 data) in grades six and seven for mathematics MAP scores and grade five for science MAP scores, comparing students of ITQG teachers with students of non-participating teachers (Henry, Murray & Hogrebe, 2012; Henry, Murray & Hogrebe, 2013).

No statistically significant differences were found for any grade in Cycle 10, although grades six and seven again showed results close to significance. More data from the Cycle 9

(2012) and Cycle 10 (2013) program evaluation reports are available from the authors' web site at http://www.mahenryconsulting.com/supportmaterials.aspx.

A closer look at the ITQG treatment teachers with two years of participation in the program was desired, in order to suggest if sustained engagement made a difference in students' achievement. The evaluators were able to determine teachers' ITQG experience over the time considered because of the data already in hand and their own evaluation records for participants.

Data available for treatment teachers with two years of ITQG participation for the years 2009 through 2012 were analyzed qualitatively by sorting and determining the mean test scores of these treatment teachers' students by year. It is recognized that some of these teachers may have been participants in previous ITQG cycles or other types of mathematics and science professional development. However, participation in professional development is common among teachers and previous ITQG participation cannot be confirmed because of limitations on available data from previous cycles. This analysis therefore assumes similar experience for all teachers prior to ITQG participation, along with two years of ITQG treatment.

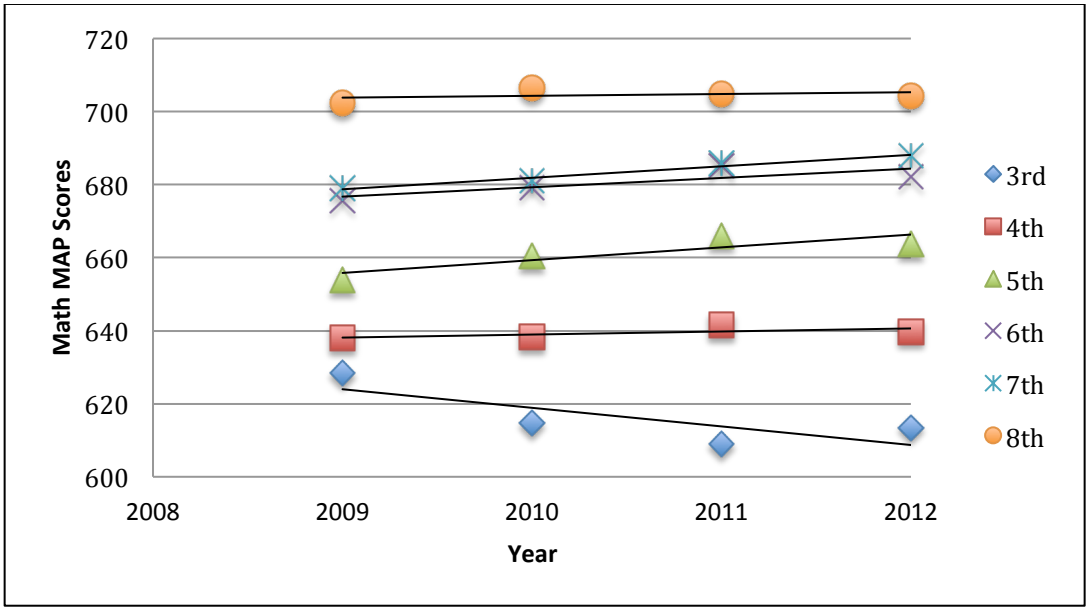
Results

The results of the analyses show that, while Cycle 10 (2012) scores tended to decline from the previous cycle, the trends for most grades evidence an overall increase (Figures 1-3). Exceptions to this pattern are noted in the discussion.

Beginning with the elementary MAP scores (Figure 1), a slight upward trend is noted for all grades except third grade. Third grade, it may be noted, had a relatively low N, which may have affected apparent trends. Although these results do not show statistically

significant change over time, the tendency for achievement to increase through most of the years examined may indicate change not statistically apparent after two years of confirmed treatment in the ITQG program.

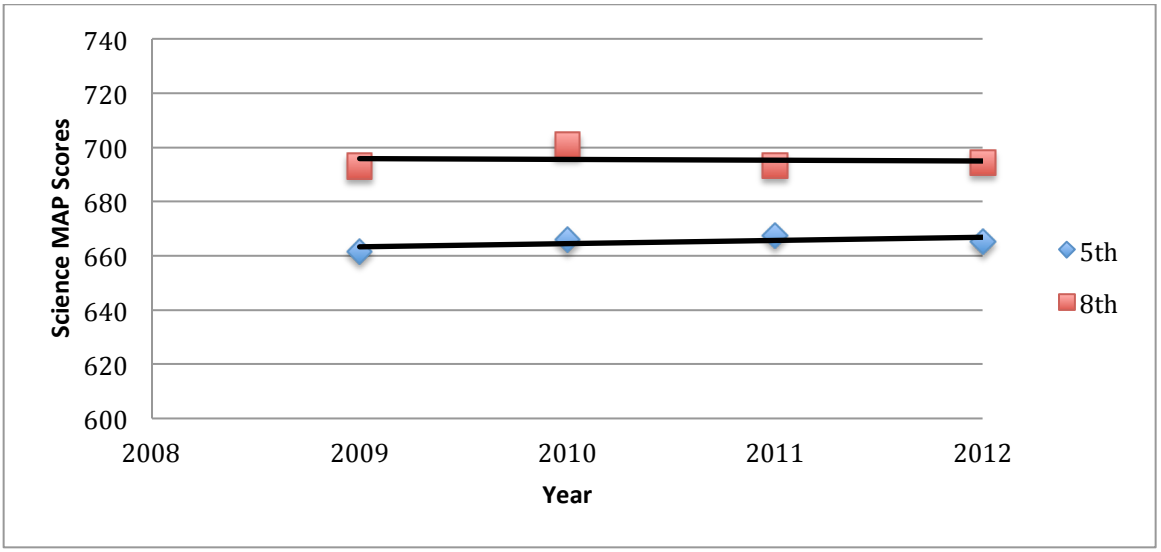
Figure 1. Trends in Mean Mathematics Achievement among Students of ITQG Teachers, 2009 - 2012, for Elementary and Middle School Grades Tested in Mathematics, Missouri Assessment Program Test Scores



3rd grade N = 412; 4th grade N = 615; 5th grade N = 2044; 6th grade N = 3,972; 7th grade N = 4,194; 8th grade N = 3,131

For the two grades tested in elementary science, a slight upward movement can be seen in the trend line for fifth-grade science. The trend for eighth-grade science has remained relatively stable since 2010 (Figure 2).

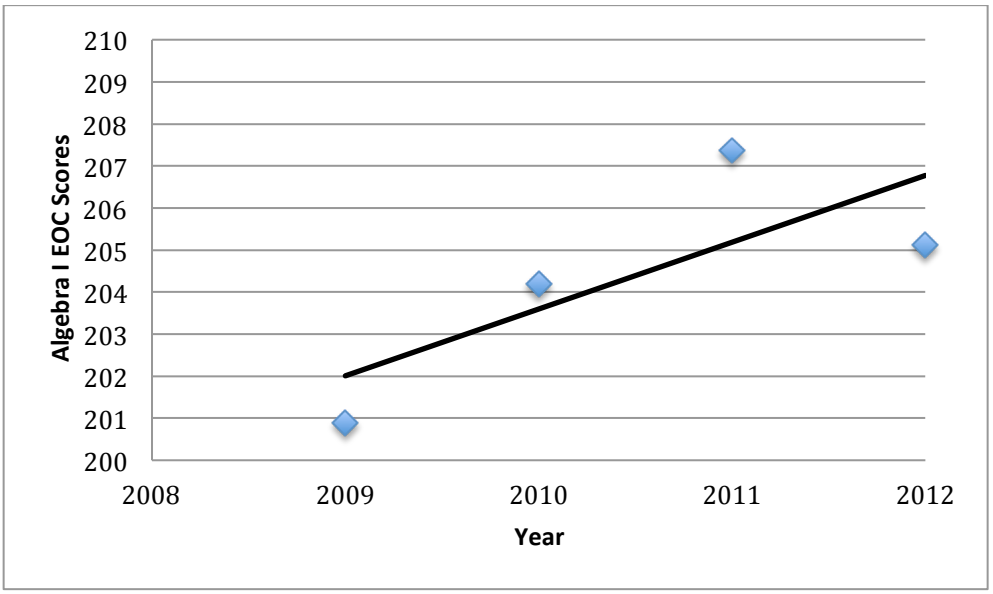
Figure 2. Trends in Mean Science Achievement among Students of ITQG Teachers, 2009-2012, for Grades Five and Eight, Missouri Assessment Program Test Scores



5th grade N = 2,044; 8th grade N = 3,127

End-of-course (EOC) tests are administered to students completing Biology I and Algebra I, after course completion, usually in high school. No ITQG project in the two years considered focused on high school science. Algebra I EOC results can be found in Figure 3.

Figure 3. Trends in Mean End-of-Course Algebra I Achievement among Students of ITQG Teachers, 2009-2012



N = 22 teacher, 1,830 students

An increasing (upward-leaning) trend line shows improving achievement through the four years considered for students of the treatment teachers.

No statistically significant gains were achieved by students of the ITQG teachers over the time period 2009 through 2012 or when compared to a statistically selected control group. However, descriptive analysis, as depicted in the trend lines for most of the grades for which test scores were considered, indicates upward movement in the achievement of students of teachers in ITQG projects.

Discussion

Standardized test data remain, in theory, the most comprehensive, complete and robust potential source of information useful for the consideration of relevant student achievement effects from large-scale teacher professional development programs such as the ITQG. The measures and interpretation of "achievement" from these data systems, being located and asserted within the formal education system, represent a less equivocal, less ad hoc, more generalizable and more cross-project-comparable standard by which to attempt to link teacher treatment with the critical objective of subsequent student impact.

However, these large data sets present challenges for such use. Access is necessarily and appropriately controlled. Data quality issues, owing to the size and details of the reporting system and other considerations (Mohan, 2015), also exist. Managing data requests within programmatic and practical constraints is complicated. Protection of confidentiality and privacy is crucial.

Nevertheless, the potential value of these data for evaluative use, additional to other primary applications, remains largely untapped nationwide. This paper, using data from a limited period of time and applying relatively simple analytic methods, shows results that

are at least suggestive of programmatic efficacy for teachers participating in the ITQG for two years. This result was not apparent in the more rigorous analyses also applied. Neither result, certainly, offers a definitive or even definite indication of whether or not ITQG projects are achieving their objectives, and the matter of achievement is much more complex than results from standardized test scores can tell.

It is important, at a time of limited funding, constraints on time and resources, and questions about the relative effectiveness of education funding applied to one program versus another to make the most of the data available. Making the most of the data in this case means not only providing access where appropriate for bona fide evaluative efforts aligned to state and federal efforts, but applying the data in as many ways possible in order to take advantage of the enormous effort and costs involved in collecting and maintaining them.

This paper has considered one more way in which the data at hand could be used for the benefit of understanding ITQG program effects. Combined with access to additional years of data corresponding to the ongoing ITQG program, with other analytical approaches brought to bear, important answers about program impact would be possible that are not now able to be addressed.

References

Aldeman, C., Carey, K., Dillon, E., Miller, B., and Silva, E. (2011). *A measured approach to improving teacher preparation.* Education Sector Policy Brief. Retrieved from http://www.unco.edu/eii/pdf/unc_eii_pb_teacherprep_part2_FINAL_021513.pdf

Blank, R.K. & De la Alas, N. (2009). *Effects of teacher professional development on gains in student achievement: How meta analysis provides scientific evidence useful to education leaders*. Council of Chief State School Officers, Washington, D.C.

Federal Education Budget Project. (2015). *Federal programs for preK-12 teachers.* Retrieved from http://febp.newamerica.net/background-analysis/federal-programs-k-12-teachers.

Henry, M.A., Murray, K.S., & Hogrebe, M. (2012). *Impact report: Program evaluation technical report, Missouri Improving Teacher Quality Grant Program, Cycle 9.* Available with related documents at http://www.mahenryconsulting.com/supportmaterials.aspx.

Henry, M.A., Murray, K.S., & Hogrebe, M. (2013). *Impact report: Program evaluation technical report, Missouri Improving Teacher Quality Grant Program, Cycle 10.* Available with related documents at http://www.mahenryconsulting.com/supportmaterials.aspx.

Henry, M.A., Murray, K.S., & Hogrebe, M. (2014). *Impact report: Program evaluation technical report, Missouri Improving Teacher Quality Grant Program, Cycle 11*. Available with related documents at http://www.mahenryconsulting.com/supportmaterials.aspx.

Miles, K. H., Odden, A., Fermanich, M., Archibald, S. (2004). *Inside the black box of school district spending on professional development: Lessons from five urban districts*. Journal of Education Finance, 30(1), 1–26.

Mohan, Rakesh. (February 2015). *The K-12 longitudinal data system (ISEE): Evaluation report 15-03.* Boise, ID:Idaho State Legislature. Retrieved from http://www.legislature.idaho.gov/ope/publications/reports/r1503.pdf