ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 19-135

STEM Instruction Improvement Programs Improve Student Outcomes

Heather C. Hill
Harvard University

Kathleen Lynch
Brown University

Kathryn E. Gonzalez
Harvard University

Cynthia Pollard
Harvard University

How should teachers spend their STEM-focused professional learning time? To answer this question, we analyzed a recent wave of rigorous new studies of STEM instructional improvement programs. We found that programs work best when focused on building knowledge teachers can use during instruction: knowledge of the curriculum materials they will use, knowledge of content and how content can be represented for learners, and knowledge of how students learn that content. We argue that such learning opportunities improve teachers' professional knowledge and skill, potentially by supporting teachers in making more informed in-the-moment instructional decisions.

VERSION: September 2019

Suggested citation: Hill, Heather, Kathleen Lynch, Kathryn Gonzalez, and Cynthia Pollard. (2019). STEM Instruction Improvement Programs Improve Student Outcomes. (EdWorkingPaper: 19-135). Retrieved from Annenberg Institute at Brown University: http://www.edworkingpapers.com/ai19-135

STEM Instruction Improvement Programs Improve Student Outcomes

Heather C. Hill, Kathleen Lynch, Kathryn E. Gonzalez, Cynthia Pollard

Suggested Citation:

Hill, H.C., Lynch, K., Gonzalez, K, & Pollard, C. (forthcoming). *STEM Instruction Improvement Programs Improve Student Outcomes*. *Phi Delta Kappan*.

Abstract

How should teachers spend their STEM-focused professional learning time? To answer this question, we analyzed a recent wave of rigorous new studies of STEM instructional improvement programs. We found that programs work best when focused on building knowledge teachers can use during instruction: knowledge of the curriculum materials they will use, knowledge of content and how content can be represented for learners, and knowledge of how students learn that content. We argue that such learning opportunities improve teachers' professional knowledge and skill, potentially by supporting teachers in making more informed in-the-moment instructional decisions.

How should teachers spend their STEM-focused professional learning time? Recent national survey data (Banilower et al., 2018) suggests teachers of STEM (science, technology, engineering and mathematics) devote their professional learning time to studying state standards, analyzing instructional materials, deepening their understanding of content and student thinking about content, learning about assessment, and studying student data. What we don't know is the extent to which such activities improve students' academic outcomes. In a recent synthesis of the research literature on STEM instructional improvement, we set out to determine just that.

We found the programs that work best tend to focus on building knowledge teachers can use during instruction: about how to use curriculum materials, about content and how content can be represented for learners, and about how students learn that content. We argue that such learning opportunities improve teachers' professional knowledge and skill, potentially by supporting teachers in making more informed in-the-moment instructional decisions.

A Study of Studies

Conducting syntheses of the literature on instructional improvement programs is not new. Such syntheses generally have two goals: to calculate the average size of program effects on student outcomes, and to learn whether specific program characteristics predict variation in these effects. However, up until recently, attempts to perform syntheses using only rigorous studies – those that compare participating teachers to a similar set of non-participants, typically via randomization – returned too few studies to achieve the latter goal. For instance, Yoon and colleagues (2007) could find only nine ELA, math, and science studies that met federal What Works Clearinghouse (WWC) evidence standards. In STEM, Gersten et al. (2014) located only five mathematics professional development studies that met WWC standards. While these numbers would allow the meta-analysis authors to calculate the overall effect of these programs on student outcomes, neither could link specific program features to those outcomes.

However, following calls for stronger research into the impact of educational interventions (e.g., Shavelson & Towne, 2001), federal research portfolios began in the early 2000s to prioritize research methods that allow for causal inference, and to require improved student outcomes as an indicator of program success. Dollars' and scholars' turn toward using causal methods and examining student-level impacts has resulted in a wealth of new studies. These new studies, we reasoned, would permit more extensive empirical analyses than prior syntheses.

With this in mind, we set out to review studies on STEM instructional improvement programs. We focused on STEM because of its importance to the national economy (Atkinson & Mayo, 2010; National Commission on Mathematics and Science Teaching, 2000), and because we wanted to provide evidence on STEM-specific professional learning activities, rather than more general activities. Because many STEM instructional improvement programs were developed in response to calls to increase both student understanding of core disciplinary ideas and student engagement with key disciplinary practices such as inquiry, argumentation and proof (NGA, 2010; NCTM, 2000; NRC 2012), we reasoned that such practices would not necessarily carry across disciplines.

As is typical in a review of this kind, we conducted extensive database searches, combed through older research syntheses, and contacted the principal investigators of studies with unpublished findings. We required studies to either feature random assignment of teachers or schools to the new program or a control group, or to have a matched comparison group of teachers identified before data collection began (for details, see Lynch, Hill, Gonzalez & Pollard, 2019).

In the end, we located 89 studies of programs that contained professional development for teachers; of these, 71 also contained new curriculum materials for teachers to use in classrooms, suggesting that program developers often paired professional development with new classroom materials. Because many consider curriculum materials to be an important source of teacher professional learning in and of themselves (Ball & Cohen, 1996), we also included six studies of new curriculum materials but no associated professional development; another three studies compared curriculum with and without professional development and were included in the count above. We located 95 studies in total.

We then read through these studies and created a dataset that contained several pieces of information from each study.

- Program type: Focus on professional development, curriculum materials, or both.
- Assessment type: Student outcomes measured via state or district standardized test, other standardized test, or researcher-designed assessment.

For teacher professional development, we further coded several other features:

- The *length* of the PD;
- The *focus* (or foci), including topics like improving teacher content knowledge, integrating technology into the classroom, and learning how to use curriculum materials;
- The *activities* that teachers engaged in during the professional development, such as reviewing student work, observing a demonstration of instruction, or working through student curriculum materials;
- The *format* of the professional development, for instance whether it was delivered during a summer workshop, contained coaching, or involved online learning.

For curriculum materials, we coded for the fraction of the original curriculum it was to replace, whether there was any background information about content or student learning embedded in the materials, and whether the materials contained kits for student experiments and activities.

Following coding, we set out to calculate an average effect size on student test scores across programs, and to understand whether any program characteristics predicted stronger or weaker student outcomes. We used a statistical technique called meta-analysis to do so; we describe the details of this analysis in Lynch, Hill, Gonzalez and Pollard (2019).

Professional Learning Programs Improve Student Outcomes

Figure 1 shows the average effect size across all STEM instructional improvement programs, both overall and broken down by the type of assessment used in the evaluation. Across all 95 studies, the average program produces a +8-percentile difference in the rank of the average treatment- and control-group student. This effect is much larger (+14 percentiles) for researcher-designed assessments than for state-standardized and other standardized assessments (+2 percentiles and +3 percentile differences, respectively). For all assessment types, our analysis shows that the difference between the treatment and control groups is not likely to be zero – meaning these effects are statistically significant, though in the case of standardized assessments, not large.

A critical question for our analysis was whether some program features outperformed others, in terms of boosting student outcomes. Figure 2 shows that programs post stronger average effects when they feature professional development paired with new curriculum materials (an average +10 percentile rank difference between treatment and control) as opposed to featuring either curriculum materials or professional development alone (an average +6 percentile-rank difference between treatment and control).

Figure 3, which shows the average impacts associated with different professional development program foci, explains this finding in more depth. Here, we compared program with specific characteristics to see whether any led to stronger student outcomes than programs without these characteristics. We also descriptively compared the impacts of programs with these characteristics relative to the average program impact in the 95-study dataset. Two program foci – helping teachers learn how to use curriculum materials, and improving teachers' content knowledge, pedagogical content knowledge, and knowledge of student learning – posted better student outcomes (of two and three percentile points better than the average program, respectively) than programs without these foci. This suggests that the combined effect of curriculum and professional development in Figure 2 may result from teachers learning how to use the materials and improving their content knowledge and knowledge of students along the way. This hypothesis is supported by a more qualitative read of the included studies; many that featured both curriculum materials and professional development often engaged teachers in solving mathematics problems, taking part in scientific investigation, watching facilitators model instruction, and studying student work.

Relationships between other foci and student outcomes were still positive, but our analysis did not indicate they were different, on average, than programs that did not include these features.

In the case of integrating technology and content-specific formative assessment, which posted stronger absolute gains than other foci, this inability to differentiate the effect from zero was due to the small number of studies included in these categories.

Figure 4 shows average impacts associated with different professional development program formats. Three formats – same-school collaboration (+2 percentiles), implementation meetings (+4 percentiles), and summer workshops (+2 percentiles)– yielded stronger gains on student assessments than programs without these formats. Same-school collaboration occurred when teachers participated in the professional development session alongside other teachers in their school, and implementation meetings allowed teachers to convene briefly with other activity participants after the start of implementation to discuss obstacles and aids to putting the program into practice. Professional learning with an online component yielded lower impacts on student learning (-4 percentiles) than programs that were entirely face to face. And programs with coaching, a popular approach to instruction improvement in many districts, yielded impacts similar to programs without coaching. However, few programs focused on extended 1:1 coaching; instead, coaching appeared more as an add-on to traditional professional development.

Our analyses detected no positive or negative associations between the activities teachers engaged during professional development and the size of program effects; this included activities such as reviewing student work, solving problems, developing curriculum materials, and reviewing both generic and their own students' work. The same was true for features of new curriculum materials – no features of those materials outperformed others.

Finally, the duration of the professional development was unrelated to student outcomes. Because this was contrary to our expectation, so we conducted several analyses to assess possible threshold effects (e.g., more than 10 hours) or a curvilinear relationship (e.g., professional development is maximally effective when between 20 and 40 hours), but we found no evidence for either.

After reading 95 studies describing STEM instructional improvement programs, what can we say?

The characteristics significantly associated with above-average student gains included:

- Professional development focused on new curriculum materials;
- Programs aimed at improving teachers' content/pedagogical content knowledge, or understanding of how students learn;
- Programs containing specific formats, including:
    - Meetings to troubleshoot and discuss classroom implementation of the program;
    - Summer workshops that allow concentrated learning time;
    - Same-school participation and collaboration.

Programs with only some or few of these characteristics may still have positive effects; however, when programs did include these characteristics, student outcomes were improved above the average program effect.

We believe that the results of this meta-analysis highlight the importance of professional knowledge for teaching. With other scholars (Ball, Thames & Phelps 2008; Shulman, 1986; Lampert, 2001), we view professional knowledge in teaching as knowing how the content, student thinking, and curriculum come together, and then making good instructional decisions based on the particulars of the situation. Programs that outperform others in our analysis tended to focus on growing this form of knowledge as opposed to general pedagogical knowledge, or more peripheral topics like technology.

What can't we say?

Meta-analyses have the advantage of examining programs implemented across a wide variety of contexts, providing some robustness to findings. However, this meta-analysis is limited in what it can say about professional learning systems "on the ground" in U.S. schools and districts. To start, each program we examined was implemented in a specific context, or a small set of contexts; whether the program would succeed in another context is an open question. Critically, an analysis found a slight trend toward smaller impacts of these programs in high-poverty settings, suggesting that interventions may work better, on average, in districts serving more advantaged students. However, we found no further interactions by student race, ethnicity, district type (urban, suburban or rural), and size of the treatment group. That said, other aspects of district and school context may interact with program efficacy. We know from studies of policy implementation, for instance, that leadership and peer support matter quite a bit in encouraging teacher take-up and use of new instructional practices (e.g., Matsumura, Garnier, & Resnick, 2010; Wanless, Patton, Rimm-Kaufman & Deutsch 2013), and the presence of competing instructional guidance and initiatives (e.g., instructional pacing guides, conflicting advice on what and how to teach) tends to dampen teacher change (Hill, Corey & Jacob, 2018). The studies we reviewed contained no information about these factors, however, so we could not test them formally in our models.

Second, the programs described here tended to be small, intensive, enrolled volunteer teachers, and were often led by university academics or researchers. By contrast, local professional development can contain myriad different offerings, with teachers spreading their time across several different settings (summer workshops, grade-level team meetings), topics (ELA, mathematics), and in sessions led by other teachers or school or district leaders. In some systems, teachers have at least partial choice over the professional development they engage, while in other systems they have very little.

One implication of these differences between the study sample and typical practice is that we don't yet know whether the features that "work" in our analysis will work in typical U.S. schools. Does a program focus on content and pedagogical knowledge improve student outcomes? Does working through new curriculum materials yield benefits to the average teacher? A lot depends upon the quality of local implementation, and in our estimation, how deeply teachers engage with the subject matter and attempt to improve their craft.

STEM-focused professional learning in wider context

As noted in the introduction, teachers of STEM engage in a wide variety of professional learning activities, often in a single year. This leads to an important question for districts: how to make more time for the kinds of learning opportunities that posted better gains? Teachers already report feeling overwhelmed by the sheer volume of reform and ever-increasing instructional responsibilities (AFT, 2017; Valli & Beuse, 2007), and it's likely that scaling back or eliminating an activity will be necessary to make room for more efficacious forms of professional learning.

Based on evidence in the literature, we would nominate eliminating "data team meetings," [1] where teachers study student data in hopes of individualizing and improving instruction. In a review separate from this one, nine evaluations of programs that featured teachers' study of data produced only two positive (and one negative) results likely to be statistically different from zero, out of a total 19 impact analyses relating program participation to student test score outcomes. As well, qualitative research suggest that teachers studying data does not itself lead to new or improved instructional techniques (Barmore, 2018; Goertz, Oláh, L. N., & Riggan, 2009), and our own observations of and participation in school-based data teams suggests that team discussions often ascribe poor student performance to factors other than instruction itself (e.g., a lack of background knowledge, a bad week, trouble at home) (see also Goertz et al., 2009). Yet recent national surveys of local professional development suggest that schools have made large investments in having teachers study student assessment data (e.g., Banilower, et al., 2018). Repurposing these meetings toward building expertise in curriculum materials and content seems natural; we caution, however, that districts will have to do so carefully, using routines and structures that focus attention squarely and in depth on instruction.

Conclusion

That these STEM instructional improvement programs boost student outcomes should be a reason for optimism among policymakers and leaders. Our findings may, for instance, help shape how states and districts choose to spend Title II dollars, funds aimed at improving teacher quality. They also suggest how leaders may narrow the scope of teacher professional learning in ways likely to increase the efficiency and impact of those efforts. Finally, our findings also suggest the importance of teachers' professional knowledge to student outcomes, a hypotheses supported by studies linking teacher knowledge to student outcomes (e.g., Baumert, 2010; Hill, Rowan & Ball, 2005) and that has implications for teacher hiring and retention.

---

[1] We differentiate "data use" programs from "formative assessment" programs in that the former typically uses data from interim or benchmark tests while the latter helps teachers create their own assessments, either from item banks, curricular assessments, or based on principles of good assessment. Programs that featured content-specific formative assessment were included in our analysis.

References

American Federal of Teachers (2017). 2017 educator quality of life survey. Washington, DC: AFT.

Atkinson, Robert D. and Mayo, Merrilea Joyce, Refueling the U.S. Innovation Economy: Fresh Approaches to Science, Technology, Engineering and Mathematics (STEM) Education (December 9, 2010). The Information Technology & Innovation Foundation, Forthcoming. Available at SSRN: https://ssrn.com/abstract=1722822

Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is—or might be—the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*, *25*(9), 6-14.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special. *Journal of Teacher Education*, *59*(5), 389-407.

Banilower, E. R., Smith, P. S., Malzahn, K. A., Plumley, C. L., Gordon, E. M., & Hayes, M. L. (2018). Report of the 2018 NSSME+. Chapel Hill, NC: Horizon Research, Inc.

Barmore, J. M. (2018). Journey from Data into Instruction: How Teacher Teams Engage in Data-Driven Inquiry. Cambridge, MA: Harvard University.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., ... & Tsai, Y. M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*(1), 133-180.

Gersten, R., Taylor, M. J., Keys, T. D., Rolfhus, E., & Newman-Gonchar, R. (2014). *Summary of research on the effectiveness of math professional development approaches.* (REL 2014–010). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from http://ies.ed.gov/ncee/edlabs

Goertz, M. E., Oláh, L. N., & Riggan, M. (2009). From testing to teaching: The use of interim assessments in classroom instruction. Philadelphia, PA: University of Pennsylvania.

Hill, H. C., Corey, D. L., & Jacob, R. T. (2018). Dividing by Zero: Exploring Null Results in a Mathematics Professional Development Program. *Teachers College Record*, *120*(6), n6.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371-406.

Lampert, M. (2001). *Teaching problems and the problems of teaching*. Yale University Press.

Lynch, K., Hill, H. C., Gonzalez, K. E., & Pollard, C. (2019). Strengthening the Research Base that Informs STEM Instructional Improvement Efforts: A Meta-Analysis. *Educational Evaluation and Policy Analysis*, 0162373719849044.

Matsumura, L. C., Garnier, H. E., & Resnick, L.B. (2010). Implementing literacy coaching: The role of school social resources. *Educational Evaluation and Policy Analysis, 32*(2), 249-272.

National Commission on Mathematics and Science Teaching for the 21st Century (US), & Chair Glenn. (2000). *Before it's too late: A report to the nation from the National Commission on Mathematics and Science Teaching for the 21st Century*. United States. U.S. Department of Education, Washington, D.C. Retrieved from https://files.eric.ed.gov/fulltext/ED441705.pdf

National Council of Teachers of Mathematics. (2000) Principles and Standards for School Mathematics. Reston, VA: NCTM

National Governors Association, Council of Chief State School Officers. (2010). *Common Core State Standards*. Washington, DC: Author.

National Research Council, Board on Science Education. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas.* Washington, DC: The National Academies Press. Retrieved from http://www.nap.edu/catalog.php?record_id=13165

Riener, C., & Willingham, D. (2010). The myth of learning styles. *Change: The magazine of higher learning*, *42*(5), 32-35.

Shavelson, R. J., & Towne, L. (Eds.). (2001). *Scientific research in education*. Washington, DC: The National Academies Press.

Wanless, S. B., Patton, C. L., Rimm-Kaufman, S. E., & Deutsch, N. L. (2013). Setting-level influences on implementation of the Responsive Classroom approach. *Prevention Science*, *14*(1), 40-51.

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007–No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from http://ies.ed.gov/ncee/edlabs

Figure 1

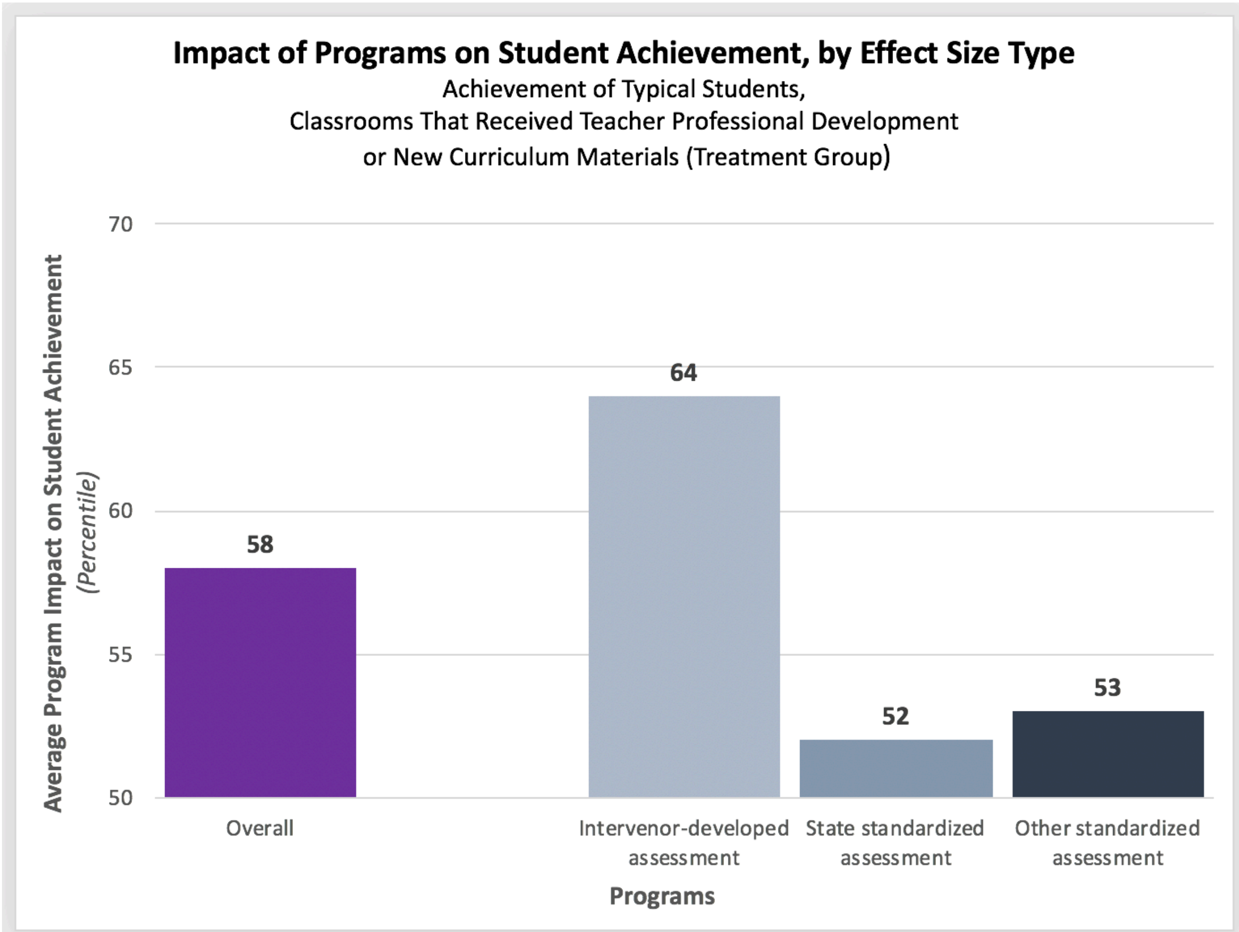
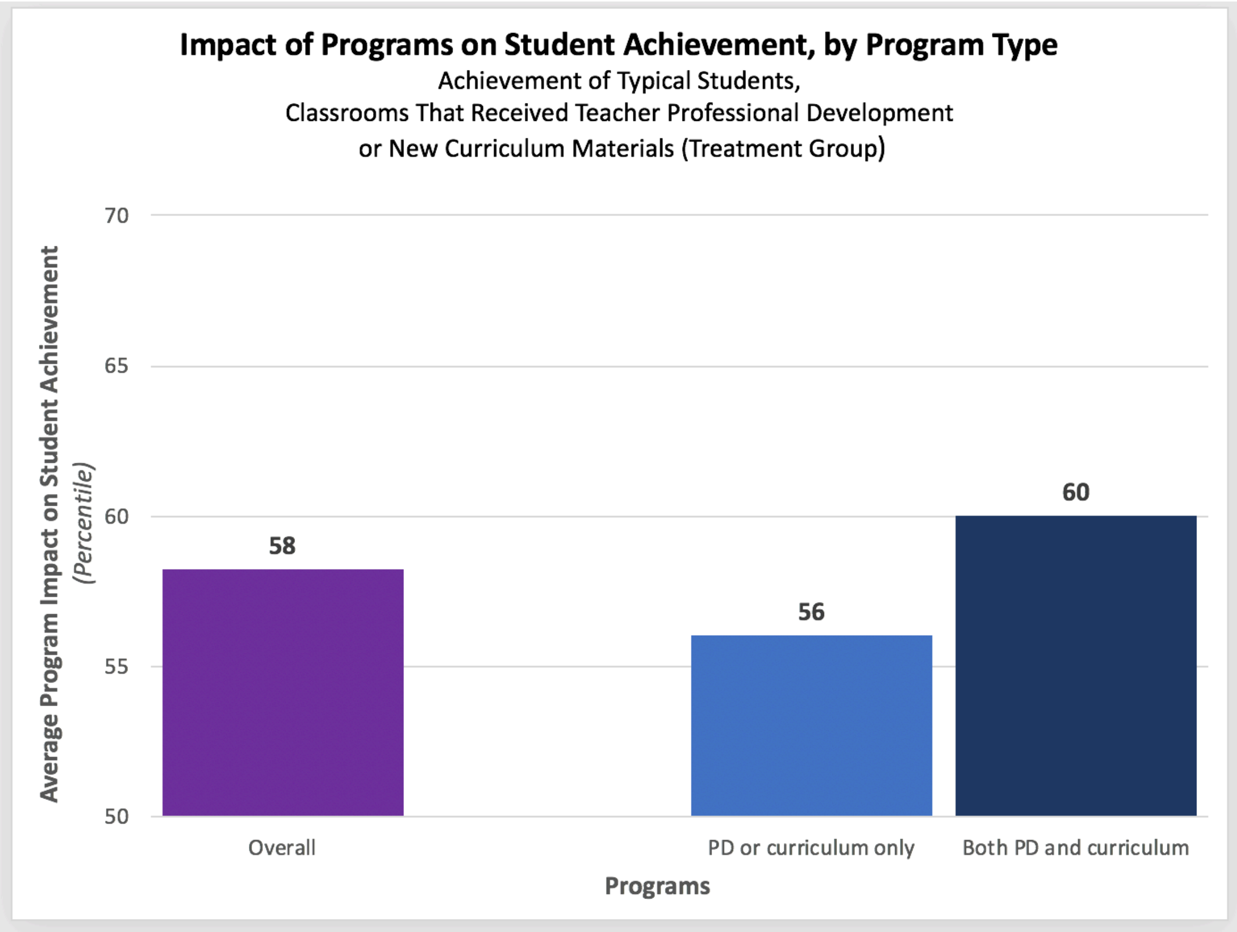

Impact of Programs on Student Achievement, by Effect Size Type
Achievement of Typical Students,
Classrooms That Received Teacher Professional Development
or New Curriculum Materials (Treatment Group)
Average Program Impact on Student Achievement
(Percentile)
70
65
60
55
50
58
64
52
53
Overall
Intervenor-developed assessment
State standardized assessment
Other standardized assessment
Programs

Figure 2



Impact of Programs on Student Achievement, by Program Type
Achievement of Typical Students,
Classrooms That Received Teacher Professional Development
or New Curriculum Materials (Treatment Group)
Average Program Impact on Student Achievement
(Percentile)
70
65
60
55
50
58
56
60
Overall
PD or curriculum only
Both PD and curriculum
Programs

Figure 3

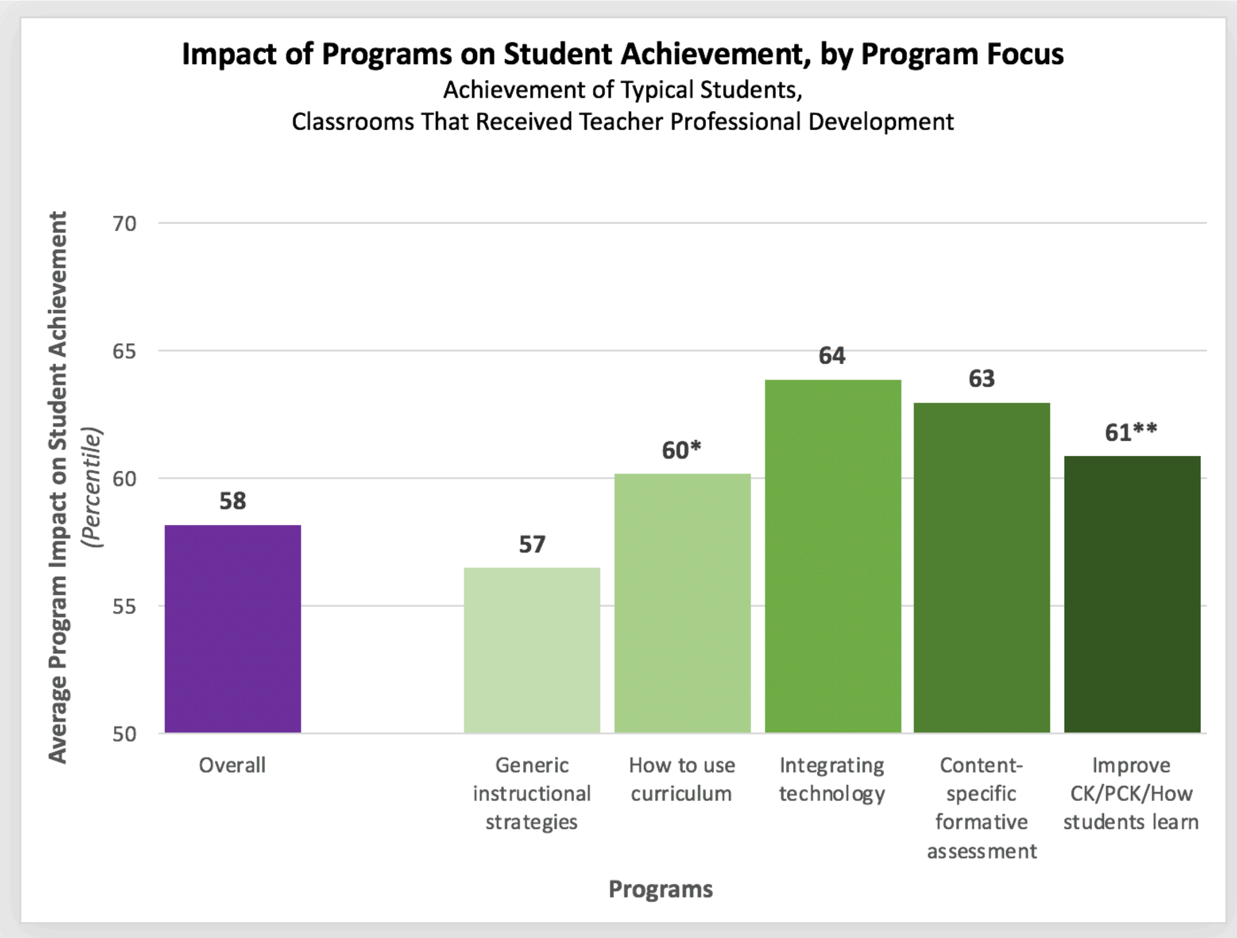
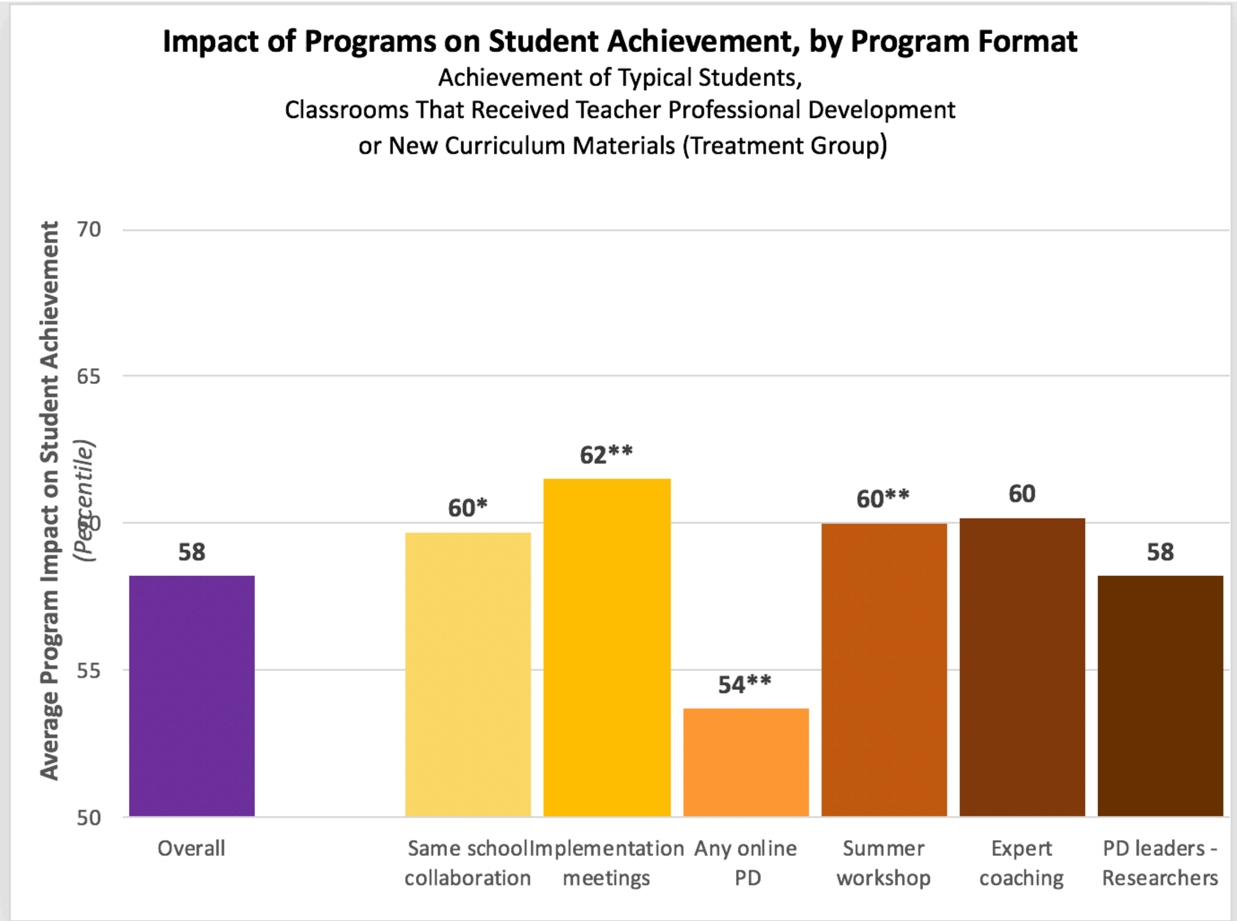

Impact of Programs on Student Achievement, by Program Focus
Achievement of Typical Students,
Classrooms That Received Teacher Professional Development
Average Program Impact on Student Achievement
(Percentile)
70
65
60
55
50
58
57
60*
64
63
61**
Overall
Generic instructional strategies
How to use curriculum
Integrating technology
Content-specific formative assessment
Improve CK/PCK/How students learn
Programs

Figure 4


Impact of Programs on Student Achievement, by Program Format
Achievement of Typical Students,
Classrooms That Received Teacher Professional Development
or New Curriculum Materials (Treatment Group)
Average Program Impact on Student Achievement
(Percentile)
70
65
60
55
50
58
60*
62**
54**
60**
60
58
Overall
Same school collaboration
Implementation meetings
Any online PD
Summer workshop
Expert coaching
PD leaders - Researchers