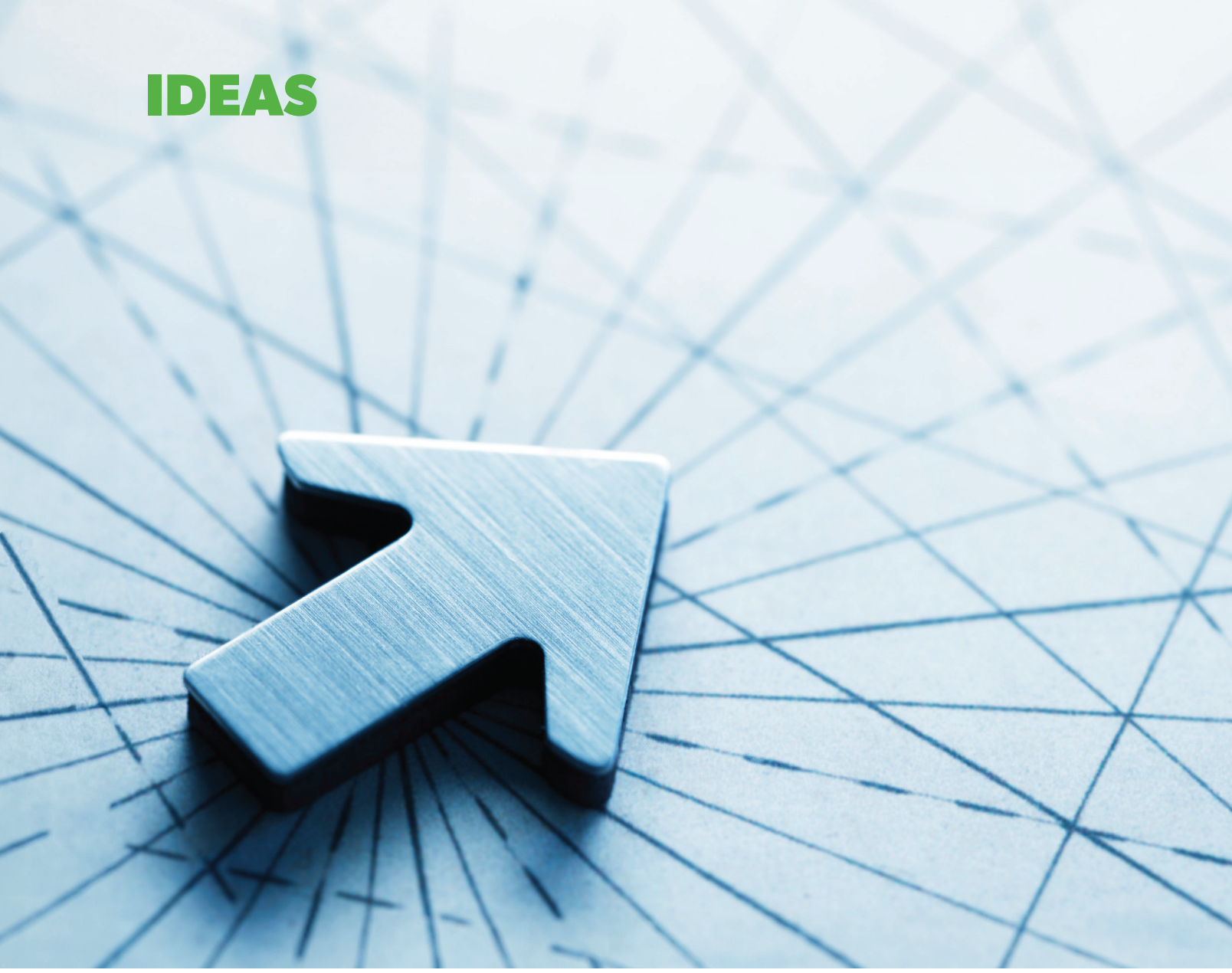


Practical evaluation points the way toward impact

BY REBECCA TAYLOR-PERRYMAN, ARIANA AUDISIO, AND LAURA MEILI

As U.S. school leaders anticipate the end of Elementary and Secondary School Emergency Relief funds and contemplate possible budget shortfalls, they will have to make hard choices about how best to leverage limited resources to improve student outcomes. With the stakes high, lawmakers and experts urge school system leaders to rely on research and program evaluation to guide decisions. Unfortunately, high-quality research that demonstrates positive outcomes and aligns with the Every Student Succeeds Act (ESSA) evidence standards (U.S. Department of Education, 2023) is rare for many types of educational interventions, including professional learning.

Evaluations by external researchers are expensive and often take multiple years to complete.

For small organizations or districts, the investment may not be feasible. Even where studies of educational interventions do exist, common challenges stand in the way of providing sufficient evidence of positive, significant results for students (Boulay et al., 2018).

Too often, ratings of professional learning are based solely on teacher satisfaction rather than Thomas Guskey's four other levels of evaluation, which examine effects on teacher practices and impact on student achievement (see p. 28 of this issue). When professional learning providers or school leaders do evaluate these other levels, they often rely on data focusing on adult perspectives or actions (e.g., teacher engagement or knowledge) and stop short of identifying whether the program led to desired changes in student achievement (Roth et al., 2019).

In addition, they often lack access to equivalent comparison groups and longitudinal data that examine effects over time. Without these factors, analyses can misattribute changes in outcomes that were already underway to a particular program.

Rather than throw up our hands and only evaluate impact every few years in a small subset of cases, which may not tell us if a particular intervention will work in other real-world contexts, we recommend that professional learning providers and school systems engage regularly in *practical evaluation*.

Educators may be familiar with the concept of practical measurement, an approach for collecting data that are useful, easy to obtain, and yet consequential — i.e., their analysis will yield meaningful insights to support improvement (Hirschboeck & Takahashi, n.d.). Practical evaluation is similar in that it is timely, uses accessible data collection methods, produces results that are easy for stakeholders to understand, and yet uses methods that go beyond pre-post comparisons so stakeholders can make inferences about causality — that is, what contribution a particular intervention may have had to an outcome.

Practical evaluation is timely, uses accessible data collection methods, produces results that are easy for stakeholders to understand, and yet uses methods that go beyond pre-post comparisons so stakeholders can make inferences about what contribution an intervention may have had to an outcome.

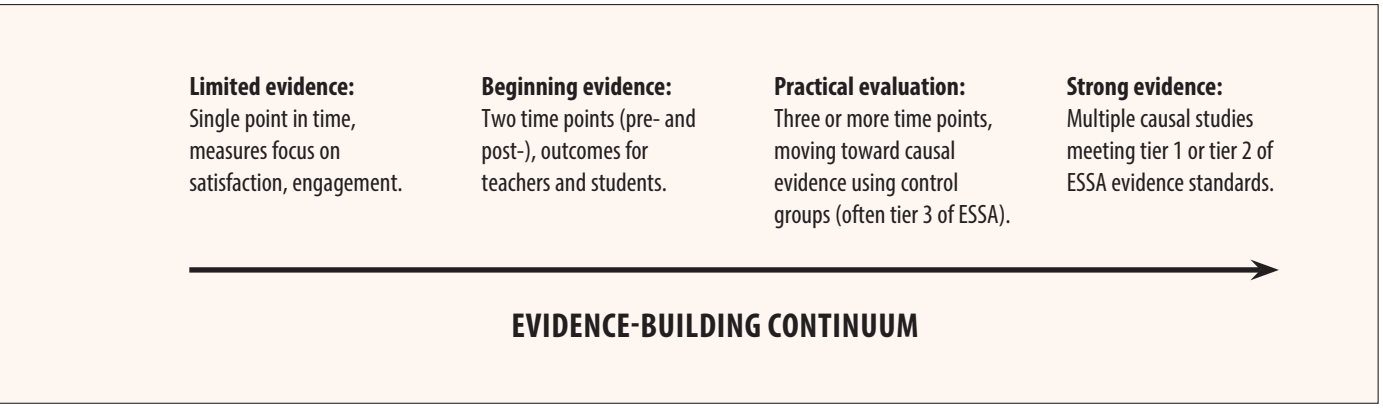
Inferences about causality are not the same as *proof* of causality but, over time, many practical evaluations can help build a stronger understanding of when and where certain interventions are likely to yield results. Through this, we can engage in the kind of continual evidence-building called for by ESSA evidence standards (U.S. Department of Education, 2023).

Building on the call from the Department of Education to grow our collective knowledge about effective innovations and on Guskey's evaluation framework for professional learning, we have developed the following evidence-building continuum (see figure on p. 44):

- With **limited evidence**, a district or school leader may only be able to answer evaluation questions such as: Did the teachers like the professional development? Did they attend?
- With **beginning evidence**, a leader can begin to understand whether instructional practice and/or student outcomes are changing, but without insights as to what may have caused those changes or how those changes relate to other trends.
- With **practical evaluation**, leaders can better understand how impacts on teaching and learning are likely related to different investments and, consequently, how to better invest time and resources in the future.
- Ultimately, **strong evidence** allows leaders to be very confident in long-term investments in programs that have consistently demonstrated impact over time.

Our organization, Leading Educators, is proud of schools and districts who have partnered with us to achieve significant effects on student learning with studies that meet the most rigorous levels of ESSA evidence standards (Audisio et al., 2023, Mihaly et al., 2022). However, we know that conditions for these studies are not always possible.

As an alternative, we also often conduct practical evaluations. Two examples of such evaluations



are presented here to illustrate practical evaluation's usefulness for understanding impact and likely reasons for the impact. In addition, we share a list of questions any district leader can ask a professional learning partner or service provider to evaluate their approach to analyzing data and the quality of evidence they share about their work.

COMPARING ACHIEVEMENT TRENDS OVER TIME IN SOUTH CAROLINA

A midsize district in South Carolina identified a set of schools performing in the bottom 5% of schools in the state and sought a professional learning partner to advance opportunity for those schools using a research-based model for school turnarounds. Over three years, we partnered with the district team to design and deliver a comprehensive set of supports for teachers, teacher leaders, principals, and district leaders aligned to new high-quality instructional materials in English language arts and mathematics.

The district's goals were to empower staff in these chronically underperforming schools to support all students with relevant, grade-appropriate lessons and close achievement gaps. Teachers, teacher leaders, principals, and district leaders engaged in ongoing coaching and professional learning facilitated by Leading Educators.

In their schools, teacher leaders led professional learning communities to guide teachers through making instructional decisions with a deeper understanding of content standards, features of high-quality curriculum, pedagogical moves that support classroom environments, and data that can inform decisions to reach rigorous, grade-level student goals.

Although we would have liked to conduct a rigorous causal study of the intervention, it was not the right fit. A causal research study requires experimental and control groups to either have similar characteristics, especially on the outcome variables, or to have similar outcome trajectories before the intervention starts. But because district leaders were understandably focused on the urgency of immediately supporting all of the turnaround schools in the district, there was likely no suitable control group that did not receive the intervention.

Nonetheless, district leaders were still eager to collect and analyze data to understand the impact of this work. We knew that if we relied on a pre-post analysis alone, we might have found positive outcomes, but we wouldn't have been able to attribute them to professional learning because other factors, such as the new curriculum or other districtwide policies, may have been responsible for the growth.

We decided on a practical evaluation approach that drew on seven years of data for the entire district to better understand trends over time in both the intervention schools and other schools in the district. We compared intervention schools' results to those of other, more advantaged schools in the district, specifically focusing on achievement gaps.

We wanted to understand whether the achievement gap between more and less advantaged schools was reduced after the intervention and whether intervention schools were able to outpace the growth of other schools. To increase confidence in the findings, we implemented statistical strategies to help us compare the change in the supported schools with the most equivalent comparison group possible within the district schools.

In the years before the partnership began, summative state assessment scores at the district's turnaround schools were *declining* by four to five points per year, while all other schools in the district *increased* by seven to eight points per year. After the partnership with Leading Educators, the average yearly growth for turnaround schools

DATA ANALYSIS STRATEGIES TO INCREASE RIGOR

We used difference-in-difference and event study strategies to control for observed and unobserved differences.

Difference-in-difference is a statistical technique that attempts to simulate an experimental research design using observational data to estimate the difference in the outcomes of a treatment and a control group after an intervention.

Our analyses included controls for percentage of students in poverty, percentage of multilingual learners, percentage of white students, percentage of students with disabilities, and grade-year and school fixed effects. See Angrist and Pischke (2009) for additional details.

not only improved, but also matched and doubled the district average growth in English language arts and math, respectively.

- In English language arts, turnaround schools and comparison schools both achieved growth of 21 points per year.
- In math, turnaround schools achieved nearly double the growth rate of other schools, at 13 points per year compared with seven points per year.

It is worth noting that these impressive results occurred in 2022 and 2023, years when achievement for the highest-need students declined nationally (National Center for Education Statistics, 2023).

But were these changes all caused by the professional learning? Some of these changes could have been caused by other districtwide initiatives. To find out, we turned to the more rigorous statistical methods. Results were as follows:

- There were positive and statistically significant improvements in English language arts that could be attributed to the professional learning because there was a comparison group sufficiently equivalent to the treatment group.
- In math, the comparison group did not show sufficient equivalence to the treatment group, which makes it harder to draw conclusions about the cause of the change.

Examining trends over a long time period (seven years) for the intervention and nonintervention schools was helpful because it allowed the district to begin to understand how the rate of change was correlated with participation, and the more rigorous statistical analyses pinpointed where we could be most confident in that correlation.

This allowed the district to more accurately identify where their investments were having the most impact and explore the root causes of those differences to guide future support for teachers and students.

COMPARING SIMILAR SCHOOLS WITH AND WITHOUT COACHING IN TEXAS

A large urban district in Texas planned to gradually roll out a new high-quality math curriculum. All schools would ultimately implement the new math curriculum, but Leading Educators partnered with the district to support an initial subset of schools.

We supported district-level instructional coaches and school leaders through coaching and professional learning sessions, and they in turn facilitated ongoing learning for teachers in their schools to support the new curriculum.

The district's goals were to ensure instruction was aligned to the instructional shifts demanded by rigorous college and career-readiness standards and for all students to gain deeper mastery of mathematical standards. Because the district's budget was limited, some schools attended professional learning sessions and received coaching support directly from Leading Educators and a different set of schools only attended Leading Educators' professional learning sessions and received coaching support from existing district coaches.

The schools included in this rollout came to participate in two different ways. All schools identified by district leaders as lowest-performing and highest need participated. In addition, district leaders were eager to support the gradual rollout with initial wins, so they allowed other schools to opt in and pilot the new math curriculum.

This had implications for our program evaluation design. Because district leaders believed it was critical to support all of the schools they saw as having the greatest need, finding a strong comparison group for those schools would have been challenging.

WHAT IS A MATCHING ANALYSIS?

Matching is a statistical technique for estimating the effect of an intervention by comparing the units that receive the intervention with units that did not receive the intervention that are similar in observed characteristics. For this analysis, we used multilevel matching with the matchMulti package in R (Pimentel et al., 2023).

However, we had other tools to analyze data. We were able to create matched comparison samples.

Because the schools came from a very large district, we could identify comparison schools that were similar to the intervention schools, thereby reducing the chances that differences we found would be due to factors like the student populations served, teachers' and leaders' knowledge and skills, and schools' motivation to participate.

With the matching analyses, we compared the change in standardized state assessment math scores for supported schools with a set of schools that were very similar in baseline outcomes and other characteristics but that did not receive the support.

Additionally, we were able to disaggregate into groups based on whether schools received coaching support directly from Leading Educators and whether they opted in or were assigned by the district.

We found that:

- All schools that received coaching from Leading Educators *grew* by 0.06 standard deviations, while matched comparison schools *decreased* by 0.04 standard deviations.
- Schools that implemented the new curriculum but did not receive Leading Educators coaching *decreased* by 0.01 standard deviations during this time. This difference suggested the importance of investing in external coaching for leaders when implementing a new curriculum.
- Schools that received Leading Educators coaching grew at equal rates regardless of whether they opted in or not. Since the lowest-performing schools who were assigned to participate started 0.6 standard deviations below the district average and the schools who opted in started roughly at the district average, this suggested this program could be effective for schools at a range of starting places.

The strong matched comparison group provided some confidence that the difference in growth was likely due to Leading Educators' coaching and not due to other factors occurring across the district at the time. Additionally, since there was no difference based on whether schools opted in, we could be more confident that the growth was likely not due to motivation to participate but instead to the intervention itself.

Observations provided additional evidence of how this growth occurred, finding improvements in use of the new, high-quality curricular materials: 78% of math materials regularly used in classrooms were considered high-quality, compared to only 23% the year before.

Nevertheless, as a practical evaluation of only one year of the initiative, the study had limitations. While the study was able to find an equivalent comparison group, the differences in rate of growth between the groups were not statistically significant, perhaps due to the size of the sample. As a result, the district was encouraged by the results but also recognized the need for additional evidence. As with all practical evaluations, repeated studies over time are needed to corroborate the findings.

CALL TO ACTION

Practical evaluations like the ones described here can be done in every systemic instructional intervention, every year, to ensure that investments have impact where it matters most: for students. Meaningful steps to increase the quality of evidence are always possible, even when conditions for more rigorous evidence standards are not met.

There is a valuable middle road between conducting rigorous, randomized evaluations and placing all our trust in single-group pre-post analyses of professional learning initiatives. With practical evaluation, we can increase the frequency with which we consider whether interventions make a difference, and whether they can do so repeatedly and in a variety of contexts. Driving improvement along the way, we can achieve greater results for all.

QUESTIONS FOR LEADERS

How can district leaders determine whether their professional learning has impact? How can partners support stronger evaluation of professional learning? The following questions can help leaders understand how trustworthy the evidence is. Answering these questions can help provide a more nuanced understanding of impact and how likely it is to be replicated.

- Are changes in outcomes measured for both teachers and students (e.g., instructional practice, student learning, or student engagement)?
- In comparison to the schools that engaged in the professional learning intervention, how did other similar schools change on the outcomes in the same time? What is similar or different about the comparison group that could have influenced those outcomes? Is there a better comparison group available?
- Is the outcome measured in a way that may leave out important information, such as only including the percentage of students performing at a particular level, which will not provide information about the movement of students above or below that threshold?
- Who is included or not included in the data analysis? For example, are some groups of students or teachers who received support excluded, and if so, why? How could that influence the results?
- What were trends like before the program started? Were schools that received the intervention already improving, and at what rate? How many years of data are included? Are any significant years left out?
- Were the changes in outcomes experienced equally by all schools who received the program?

REFERENCES

Angrist, J.D. & Pischke, J.S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Audisio, A., Taylor-Perryman, R., Tasker, T., & Steinberg, M.P. (2023). *Does teacher professional development improve student learning? Evidence from Leading Educators' fellowship model.* (EdWorkingPaper: 22-597). Annenberg Institute at Brown University. doi.org/10.26300/ah2f-z471

Boulay, B., Goodson, B., Olsen, R., McCormick, R., Darrow, C., Frye, M., ... & Sarna, M. (2018). *The Investing in Innovation Fund: Summary of 67 evaluations. Final report.* (NCEE 2018-4013). National Center for Education Evaluation and Regional Assistance.

Hirschboeck, K. & Takahashi, S. (n.d.) *What we're learning: Using practical measures to support school improvement initiatives*. Bill & Melinda Gates Foundation. usprogram.gatesfoundation.org/news-and-insights/articles/what-were-learning-using-practical-measures-to-support-school-improvement-initiatives

Mihaly, K., Opper, I.M., & Greer, L. (2022). *The impact and implementation of the Chicago Collaborative teacher professional development program*. RAND Corporation.

National Center for Education Statistics. (2023). *Reading and mathematics scores decline during COVID-19 pandemic*. www.nationsreportcard.gov/highlights/ltt/2022/#intro

Pimentel, S.D., Page, L.C., & Keele, L. (2023). *An overview of optimal multilevel matching using network flows with the matchMulti package in R**. cran.r-project.org/web/packages/matchMulti/vignettes/multiMatch_vignette.pdf

Roth, K.J., Wilson, C.D., Taylor, J.A., Stuhlsatz, M.A., & Hvidsten, C. (2019). Comparing the effects of analysis-of-practice and content-based professional development on teacher and student outcomes in science. *American Educational Research Journal, 56*(4), 1217-1253.

U.S. Department of Education. (2023). *Non-regulatory guidance: Using evidence to strengthen education investments*. Author.

•

Rebecca Taylor-Perryman (rtaylorperryman@leadingeducators.org) is the managing director of data and evaluation, Ariana Audisio (aaudisio@leadingeducators.org) is the associate director of data and evaluation, and Laura Meili (lmeili@leadingeducators.org) is the chief impact officer at Leading Educators.