education
analytics

Evaluating the Impact of COVID-19-Related Interventions

Key Considerations and Available Approaches

This resource provides guidance for education agencies that are considering how and when to evaluate programs designed to address the impacts of COVID-19 on student outcomes. We present some key questions to consider, share common data requirements, and outline different options for designing the analytic approach.

Overview

The U.S. Department of Education is funding interventions that accelerate student learning affected by COVID-19, and it has provided guidance on how best to use these funds. For the most part, local education agencies (LEAs) have broad discretion over what programs (interventions) best meet their needs, and which ones they will adopt and implement so that students do not fall further behind next school year. At the same time, LEAs will want to determine whether the programs they implement work well for all participating students, and especially for their highest-need learners. Evaluation of these efforts can ensure not only that supports are effective, but also that money provided is well spent. Finally, as one-time funding phases out, LEAs will be left to decide whether to continue these efforts in place of other existing instructional strategies and supports, or to phase COVID interventions out of their program offerings.

Given that evaluation can be a time-consuming endeavor, we lay out some questions below for LEAs to consider in preparation for this work, share guidance on data requirements for program evaluation, provide some suggestions for evaluation designs, and describe the process Education Analytics takes to partner with education agencies interested in evaluating their efforts.

What are the key questions education agencies should consider?

1 *What is the program being offered?*

EXAMPLE
A new 1:1 math tutoring program in response to COVID-19 (or one modified for COVID-19)

WHY IS THIS IMPORTANT?
This helps us know what types of measures to use for understanding impact, and whether they already exist or if we should develop new ones (e.g., questionnaires).

Key Questions to Consider

2

What are the expected or desired outcomes of participating in the program?

EXAMPLE
Improve student math achievement as measured by quarterly benchmark assessments

WHY IS THIS IMPORTANT?
Clearly stating the intended outcome informs what measures are available to assess the outcome in an evaluation (or whether new measures are needed).

3

What is the theory about how the program impacts student outcomes?

EXAMPLE
1:1 math tutoring for 10 weeks, 3 times per week will help struggling students become minimally proficient in math, thereby enabling those students to be successful in the general program.

WHY IS THIS IMPORTANT?
Articulating the theory ensures that the logic for the program is known and appropriate for the services provided. A t*heory of action* or *logic model* explicitly showing the causal links between program activities and the intended outcomes identify what outcomes and processes should be measured.

4

How do you identify students for the program?

EXAMPLE
Based on students' 2 most recent benchmark math assessments, students in the lowest math level are identified for the program.

WHY IS THIS IMPORTANT?
This is the most critical component of a successful evaluation. How students are assigned to the intervention determines how to identify a suitable comparison group for the analysis. *See the following section on understanding the importance of selection.*

Key Questions to Consider

5 *What are the key components of the program needed for it to work?*

WHY IS THIS IMPORTANT?
Identifying the components helps specify what metrics should be used to determine if the program was implemented as intended.

EXAMPLE
1:1 tutoring, trained tutors, and a quiet location to engage in tutoring

6 *When was the program implemented and for how long?*

WHY IS THIS IMPORTANT?
This allows us to target when we would measure the impact of the intervention, and what data would be needed to serve as a baseline.

EXAMPLE
Spring semester 2021

7 *For programs implemented across multiple sites (i.e., classrooms, schools, or districts), how similar is the intervention?*

WHY IS THIS IMPORTANT?
When programmatic sites differ in their implementation, we may need additional measures to contextualize those differences and fully evaluate the impact of a program. We also may need to investigate which site designs worked best, if there are substantial differences across them.

EXAMPLE
All participating sites implemented a common tutoring program with the same identification criteria. Some schools implemented the program before or after school, while others implemented the program during school.

Key Questions to Consider

8 *How are you tracking which students participate and for how long?*

EXAMPLE
A special program code was created in the student information system. Each tutoring event has been logged as an intervention session in the student information system using the special program code.

9 *Are there other measures to consider in measuring impact?*

EXAMPLE
Each site has a site coordinator for the program who completes a weekly observational report about the successes and challenges of implementing the program.

The Importance of Selection into the Program

When participation in a program is voluntary (meaning students or families can opt in or out), there are always outside factors (beyond what we can measure) that could affect both the decision to participate and the outcomes of interest. For example, students who are less academically motivated may be less likely to participate in a voluntary program, but they also may be more likely to experience lower academic achievement.

In these cases, we cannot simply compare the participating and non-participating groups of students to assess whether participating in the intervention is associated with a measurable difference in outcomes. Following the earlier example, if the less-motivated students are part of the "comparable" group of students, then we would be overestimating the impact of the program (since these students would have lower academic achievement due in part to their lower motivation, and not simply because they did not participate in the intervention). This means that evaluations that do not account for differences in these kinds of unobserved factors are likely to result in biased (i.e., incorrect) findings.

What data are needed to conduct a rigorous program evaluation?
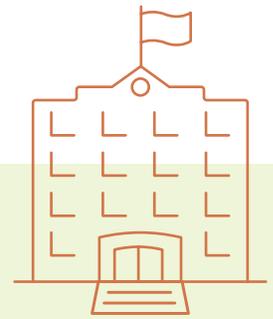
Student-Level Data

Student IDs (to be able to identify each unique student)

Demographic data, such as race/ethnicity and gender/sex

Programmatic services, such as whether students receive free or reduced-price lunch, special education services, or English language services

Outcome data related to the program, such as attendance, discipline, social-emotional learning measures, and/or interim or summative assessments used in the district or state
- This is required for both the duration of the program (to measure impact) and for previous years (to serve as a baseline)

Teacher-student course linkages (if the program is delivered primarily through an instructor)
- **Teacher IDs**
- **Course enrollment duration** (i.e., how long each student was enrolled in that course)

School-student linkages (if the program is delivered at the school level across sites)
- **School IDs**
- **School enrollment duration** (i.e., how long each student was enrolled in that school)

For other programs, **data on student participation** such as program entry/exit and program dosage (e.g., in hours or days).
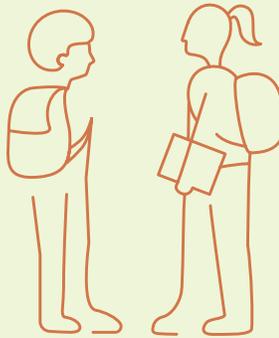
Site-Level Data

Number of sites that offered the program

Number of teachers or specialists participating

Number of students at each site

Curriculum or instructional materials used at each site (if different)

Other information describing fidelity of implementation, such as:
- Percentage of teachers and coaches trained in the curriculum
- Percentage of classrooms in which key program components were observed to be present
- Percentage of students accessing web-based materials
- Ratio of students to instructors

Time Period

Ideally, we would include student data from both **before**, **during**, and **after** the program is implemented. That way, we can determine their starting points (baseline) and how the program potentially impacted them during and after the program.

Teacher- or Specialist-Level Data

Measures of teacher/specialist practice (e.g., observational ratings data)

Measures of teacher/specialist experience (e.g., years of experience in the role or district)

What approach could you use to conduct the evaluation?

Below, we list analytic approaches that are available for evaluating impact; note that this is not meant to be an exhaustive list of all possible evaluation designs, as the design used depends on many of the factors listed above. Additionally, depending on the use of the results, readers may want to consult the **What Works Clearinghouse** (WWC) Standards for a more in-depth explanation of designs that meet federal guidelines for evaluation.

Each row in the table provides a short description of a design, **starting from the least rigorous to the most rigorous design options.** As the designs become more rigorous, they typically provide stronger evidence for a causal impact of the program or intervention, at the cost of greater complexity. Each educational institution will need to select a design that provides the appropriate trade-offs between rigor and complexity for the organization's needs. The remaining columns in the table define the comparison being made, an example of an outcome measure to measure the program impact, a justification for how that design increases rigor, and other considerations when selecting that design.

Design	Comparison	Example	Increases Rigor by	Other Considerations
Before and after change with similar students/ school comparison	Change in outcome for participating students/schools compared to change over the same period for a set of students/ schools that are similar but did not participate	Participating schools increased the proportion of proficient students by +7 percentage points, versus by 0 percentage points for the comparison schools.	Compared to analyzing a change in outcomes for only those students/schools who participated, this includes a comparison group and helps to rule out some pre-existing differences among students/schools as the cause of change.	+ If similar schools are from pre-established peer groups, no need to create groups; such comparison may be highly credible to stakeholders – Need to ensure peer groups are comparable; subjective groupings may be open to dispute – Cannot be used when all schools participate in the program
Cross-sectional observational study with statistical control (e.g., using a model that predicts the outcome based on school characteristics and program participation)	Average outcome for program students/schools compared to the outcome predicted by the statistical model	Positive and statistically significant difference in outcome for participating schools compared with non-participating schools	Controls for differences between students/schools other than program participation that are related to outcomes Provides an indicator of how likely results are due to chance	+ Helps ensure that comparisons are more objective – Requires some students/schools do not participate in the program – Requires careful thought to identify the characteristics that influence outcomes besides program participation – Requires data be collected about these characteristics – May require outside experts and specialized software to implement

Design	Comparison	Example	Increases Rigor by	Other Considerations
Difference in Differences or **Comparative Trend Analysis** (i.e., statistical analysis comparing the *change in outcome* over time for the program students/school with the *change in outcome* for a comparable group)	Compares the change over time in participating students/schools with the change over time in non-participating students/schools	Participating schools increased proficiency by 6 percentage points, while non-participating schools increased 3 percentage points.	Focusing on difference in rate of change which helps to rule out pre-program differences in outcomes as a contributor to post-program differences	+ Comparison group does not need to be highly similar to participant group credible to stakeholders − Works best with statistical controls that account for changes over time in *other* school characteristics that could influence outcomes + Adding more pre-program years to the trend allows checking whether rate of change was different between treated and comparison students/schools before program − Likely requires outside experts and specialized software to implement
Propensity score matching (PSM) (i.e., statistical analysis that finds a set of students/schools that did not participate in the program but are similar to students/schools that participated, in terms of pre-existing characteristics that influence outcomes)	Average outcome for participating students/schools compared with average outcome or change in demographically similar students/schools not participating	Participating schools gained 5% in average percent of proficient students this year versus 3% gained in demographically similar schools.	Adding comparison group that is highly similar to the participating students/schools on pre-exiting characteristics that influence outcomes	+ Meets WWC Standards + Does not require random assignment of students to program − May not be able to find matches for all participating students/schools − Can't control for unmeasured differences (e.g., self-selection of more motivated students into program) that influence outcomes
Randomized Control Trial (RCT) (i.e., randomly assign the program to students or to one group of schools across the district/state)	Compare the after-implementation outcome for the students/schools assigned to the program or intervention to the outcome for students/schools not assigned	Participating schools gained 5% in average percent proficient this year versus 3% gained in comparison schools.	Eliminates potential influence that self-selection might have on outcomes by randomly assigning students or schools to a program and then comparing results	+ Meets WWC Standards − Need to randomize assignment before program is implemented; can't be used for programs already in place − May be hard to convince program designers to randomly assign students or schools − Likely requires outside experts and specialized software to implement

What services does Education Analytics offer?

EA's typical impact evaluation services include:

Understanding the **theory of action** and constructing **logic models** with program leaders and via documentation

Data inventory of elements aligned to the theory of action/logic model

Collection of **contextual information about the program** (i.e., how it was/is being offered and any data collected on fidelity of implementation [FOI])

Interviews with program leadership and select participants to understand how the program was implemented

Analysis of **student participation and outcome data** utilizing appropriate methods, including those which meet What Works Clearinghouse (WWC) standards

Summary of findings in the form of slides and/or a report

Technical documentation of findings in the form of a written report

In some cases, our evaluation services can also include additional qualitative elements:

Development and collection of **Fidelity of Implementation** (FOI) indicators

Rigorously implemented interviews with program participants for a full qualitative evaluation

Construction and administration of **surveys** to better understand implementation and impact

Analysis of implementation fidelity and program impact on **intermediate outcomes** (though for shorter-duration programs, intermediate outcomes may not be relevant)

Research study submission to **IRB** or district research review board

EA's Impact Evaluation Process

1 Conduct a needs assessment, facilitated using questions from this document

2 Assess data availability *(See previous data section)*

3 Finalize scope and budget

Determine which EA evaluation service is preferred

Pending data availability, finalize the analysis plan for each cost driver:
- Number and type of data sources
- Type of modeling approach required
- Format of final deliverables

4 Co-build evaluation design and modeling decisions

Review programmatic material related to design, implementation, and previous external evaluations (if applicable)

Make evaluation design and modeling decisions with stakeholders

Preliminary analysis results and feedback with stakeholders
- Examine preliminary results and how they meet expectations
- Re-visit modeling decisions as needed

Final analysis production and quality control

What should my agency do next?

If your education agency has the staff and time available to conduct an evaluation of your programs aimed at addressing the impact of COVID-19, this document provides a framework to help record the key elements of the program to consider as part of the evaluation design. Taking an inventory of the data you have available and cross-referencing that with the data requirements provided in this document can help you align the outcomes you wish to measure with the variables available to do so. Finally, the list of available analytic approaches provides a starting point for ensuring the analysis you choose will answer the questions you aim to answer with the data you have available. Education Analytics can also provide direct support to your agency by consulting on the planning and design of your evaluation, conducting the statistical analysis, and summarizing actionable findings from both quantitative and qualitative evidence. To get in touch with us, please visit **edanalytics.org/contact.**

Available Funding

USDE guidance states that ARP funds can be used for the following purposes:

Providing principals and other school leaders with the resources necessary to address the needs of their individual schools.

Activities to address the unique needs of low-income children or students, students with disabilities, English learners, racial and ethnic minorities, students experiencing homelessness, and children and youth in foster care, including how outreach and service delivery will meet the needs of each population.

Addressing the academic impact of lost instructional time among an LEA's students, including low-income students, students with disabilities, English learners, racial and ethnic minorities, students experiencing homelessness, and children and youth in foster care. These measures include:

- Administering and using high-quality assessments that are valid and reliable to accurately assess students' academic progress and assist educators in meeting students' academic needs, including through differentiating instruction.

- Implementing evidence-based activities to meet the comprehensive needs of students.

- Providing information and assistance to parents and families on how they can effectively support students, including in a distance learning environment.

- Tracking student attendance and improving student engagement in distance education.