



Working Paper

SUSAN BUSH-MECENAS, JONATHAN SCHWEIG, MEGAN KUHFELD, LOUIS T. MARIANO, AND  
MELISSA KAY DILIBERTI

# Research, Interrupted

---

Addressing Practical and Methodological Challenges  
Under Turbulent Conditions

RAND working papers are intended to share researchers' latest findings and to solicit informal peer review. They have been approved for circulation by RAND Education and Labor but have not been formally edited or peer reviewed. Unless otherwise indicated, working papers can be quoted and cited without permission of the author, provided the source is clearly referred to as a working paper. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. **RAND**® is a registered trademark. Learn more at [www.rand.org](http://www.rand.org).

For more information on this publication, visit [www.rand.org/t/WRA1037-1](http://www.rand.org/t/WRA1037-1).

#### **About RAND**

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest. To learn more about RAND, visit [www.rand.org](http://www.rand.org).

#### **Research Integrity**

Our mission to help improve policy and decisionmaking through research and analysis is enabled through our core values of quality and objectivity and our unwavering commitment to the highest level of integrity and ethical behavior. To help ensure our research and analysis are rigorous, objective, and nonpartisan, we subject our research publications to a robust and exacting quality-assurance process; avoid both the appearance and reality of financial and other conflicts of interest through staff training, project screening, and a policy of mandatory disclosure; and pursue transparency in our research engagements through our commitment to the open publication of our research findings and recommendations, disclosure of the source of funding of published research, and policies to ensure intellectual independence. For more information, visit [www.rand.org/about/research-integrity](http://www.rand.org/about/research-integrity).

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

Published by the RAND Corporation, Santa Monica, Calif.

© 2023 RAND Corporation

RAND® is a registered trademark.

#### **Limited Print and Electronic Distribution Rights**

This publication and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited; linking directly to its webpage on [rand.org](http://rand.org) is encouraged. Permission is required from RAND to reproduce, or reuse in another form, any of its research products for commercial purposes. For information on reprint and reuse permissions, please visit [www.rand.org/pubs/permissions](http://www.rand.org/pubs/permissions).

## About This Working Paper

---

The COVID-19 pandemic caused tremendous upheaval in schooling. In addition to its devastating effects on students' academic development, the disruptions to schooling had important consequences for researchers conducting effectiveness studies on educational programs during this era. Given the likelihood of future large-scale disruptions, it is important for researchers to plan resilient studies and think critically about possible adaptations when such turbulence arises. In this article, we utilize qualitative case study analysis to examine how researchers evaluating educational programs in the pandemic period adjusted to turbulent conditions through design pivots to ensure the feasibility of research. We find that researchers struggled to strike a balance between the evaluations that were intended and those that could realistically be accomplished. We identify how the challenges of the pandemic period and design pivots raised potential threats to validity, illuminate some promising practices that arose during the pandemic period, and provide recommendations for future research and evaluation programs focused on studying the effectiveness of educational programs during times of profound disruption.

### RAND Education and Labor

This study was undertaken by RAND Education and Labor, a division of the RAND Corporation that conducts research on early childhood through postsecondary education programs, workforce development, and programs and policies affecting workers, entrepreneurship, and financial literacy and decisionmaking.

More information about RAND can be found at [www.rand.org](http://www.rand.org). Questions about this report should be directed to [jschweig@rand.org](mailto:jschweig@rand.org), and questions about RAND Education and Labor should be directed to [educationandlabor@rand.org](mailto:educationandlabor@rand.org).

### Funding

The research reported here was sponsored by the Institute of Education Sciences, U.S. Department of Education, through grant R305U200006 to RAND. The opinions expressed in this report are those of the authors and do not represent views of the Institute or the U.S. Department of Education. This paper and the research also benefited from the thoughtful review of Ben Master.

# Contents

---

About This Working Paper.....	iii
Figures and Tables.....	v
Chapter 1. Introduction.....	1
Practical Challenges Posed by Large-Scale School Disruptions.....	2
Methodological Challenges Posed by Large-Scale School Disruptions .....	4
The Current Study .....	7
Chapter 2. Data and Methods .....	8
Limitations.....	10
Chapter 3. Findings .....	12
Challenges To Intervention Feasibility Caused by Situational Complexity .....	12
Difficulty With Study Recruitment .....	14
Data Availability and Concerns About Data Quality .....	16
Concerns with Extrapolation and Generalization.....	18
Chapter 4. Discussion and Summary.....	20
Conclusion.....	21
Abbreviations .....	24
References .....	25

# Figures and Tables

---

## Tables

Table 1.1. Categories of Validity Threats .....4  
Table 2.1. Categories of Cases for Sampling ..... 8  
Table 2.2. Descriptions of Included Studies.....9

# Chapter 1. Introduction

---

Policymakers in the U.S. have increasingly called for education intervention and policy adoption to be guided by evidence-based research (e.g., Slavin, 2002). In 2001, the No Child Left Behind Act (U.S. Department of Education [U.S. ED], 2002) mandated that scientific evidence serve as the basis for federally funded interventions. Since that time, the Institute for Education Sciences (IES) of the U.S. ED has invested in studies that use experimental and quasi-experimental designs to provide rigorous evidence of what works (e.g., Slavin, 2008). To support this agenda, U.S. ED created the What Works Clearinghouse (WWC) and the Standards for Excellence in Education Research (SEER) principles (IES, 2022b). Collectively, WWC and SEER codify expectations that education studies should mitigate threats to validity, thereby increasing the credibility of such research. This helps ensure researchers and policymakers draw appropriate conclusions about which interventions and policies are likely to be effective in a specific school or district context (e.g., Joyce & Cartwright, 2020).

Large-scale<sup>1</sup> disruptions to schooling—including both those during the COVID-19 pandemic and at other time points—have important implications for the researchers who planned and continued to conduct research under these turbulent conditions. Atypical schooling conditions potentially jeopardize the integrity of findings from individual studies conducted under these conditions and impede efforts to advance our collective understanding of “what works” in education. Although the COVID-19 pandemic-related school closures were unprecedented in scale, large-scale disruptions to schooling were already common by 2020. Between 2011 and 2019, over 13 million students in the U.S. were impacted by prolonged unplanned school closures primarily due to extreme weather phenomena (including hurricanes), natural disasters (including wildfires), teacher strikes, environmental issues, and violence (Jahan et al., 2022). Furthermore, there is an emerging consensus that further disruptions to schooling are inevitable in the future (e.g., Inter-agency Network for Education in Emergencies [INEE] 2020; Salman, 2022). Given the likelihood of future large-scale disruptions, it is important for researchers seeking to evaluate education interventions and policies to plan resilient studies and think critically about possible adaptations when such turbulence arises. A systematic assessment of common challenges and successes can help to ensure that future studies meet standards for high-quality research (Institute of Education Sciences [IES], 2022a).

In this article, we examine the COVID-19 pandemic period as one particularly salient case of a large-scale disruption to schooling that exemplifies challenges faced by researchers evaluating education interventions and policies under turbulent conditions. The pandemic period represents

---

<sup>1</sup> We define large-scale disruptions as those that are extensive in terms of duration and far-reaching in terms of the number of students affected.

a strategic case as the disruption was profound, long lasting, and widespread, which allowed us to explore variation across study contexts. We utilize qualitative analysis of interviews with research study leaders to illuminate some promising practices that arose during the pandemic period and to provide recommendations for future education research conducted during times of profound disruption.

We draw on a growing body of research in international development about conducting research amid humanitarian crises (Bakrania et al., 2021; Hassnain et al., 2021; Nene Odjidja & Alves do Reis, 2021; Puri et al., 2017). We use this literature as a conceptual framework to guide our analysis of the pivots and adaptations made by education researchers who conducted studies during the COVID-19 pandemic. This literature has identified several practical challenges that are commonly encountered by researchers in humanitarian settings. The same literature has also identified methodological issues that threaten the “validity of an impact evaluation” (Puri et al., 2017, p. 523) conducted under these conditions. We use this international development framework and overlay concepts from a widely- referenced validity framework developed by Campbell and colleagues (Campbell & Stanley, 1963; Cook & Campbell, 1979; Shadish et al., 2002) to guide our analysis. More specifically, we frame the practical and associated methodological challenges encountered by researchers during an extreme period of school disruptions in terms of specific threats to internal validity, construct validity, statistical conclusion validity, and external validity. In remainder of this section, we describe the specific practical challenges as well as the methodological issues posing specific threats to validity due to large-scale school disruptions like the COVID-19 pandemic.

## Practical Challenges Posed by Large-Scale School Disruptions

Building on the international development literature, we defined four specific practical challenges that are distinguishing features of conducting education research in the midst of large-scale school disruptions like the COVID-19 pandemic: 1) complex settings marked by turbulence and instability; 2) difficulty in accessing research populations and collecting data; 3) lack of local capacity to support research; and 4) limited availability of good-quality data (Bakrania et al., 2021; Hassnain et al., 2021; Nene Odjidja & Alves do Reis, 2021; Puri et al., 2017).

### *Complex Settings*

Schools and districts are complex, dynamic systems (Groff 2013) marked by turbulence and instability even in normal school years (Myers 2014). Turbulence regularly arises for a variety of reasons, including changes to district capacity and priorities, organizational culture, curriculum, standards, and assessment. However, large-scale disruptions like the COVID-19 pandemic introduce additional complexities, disrupting even the most durable aspects of schooling—things like instructional mode, timing and length of the school year, student enrollments, and grade progressions. For example, as mentioned above, schools in all 50 states closed to in-person

instruction in spring 2020. They reopened to in-person instruction on different timelines and had different policies on instructional modes. In sum, the COVID-19 pandemic caused widespread breakdowns in both logistics and school infrastructure (Puri et al., 2017).

### *Difficulty in Accessing Research Populations and Collecting Data*

Education researchers often rely on direct observation, interviews, and focus groups with students and school staff—often conducted in-person at schools—in their studies. School closures presented serious difficulties for researchers looking to access research populations, as researchers were prohibited from visiting schools or interacting in person with students or school staff throughout the pandemic. Even as schools reopened for in-person instruction in the 2020–2021 school year, many state departments of public health recommended that schools continue to limit nonessential visitors (e.g., California Department of Public Health, 2022). While researchers may have increasingly pivoted to conduct their data collections virtually, this shift may have excluded particularly vulnerable populations from participating in research because of systematic inequities in home internet access (e.g., Stelitano et al., 2020).

### *Lack of Local Capacity to Support Research*

Even under ideal conditions, conducting school-based research involves the input and support of a variety of system stakeholders, including (but not limited to) district administrators, principals, teachers, and parents. At minimum, these stakeholders need to grant researchers permission to access research populations. But often these stakeholders support research in much more involved ways, including participating in implementation trainings, facilitating the distribution of consent forms, establishing data use agreements, fulfilling administrative data requests, and reviewing study procedures and protocols to ensure that they meet standards for ethical research (e.g., Bartlett et al., 2017). During the pandemic, schools were faced with serious staffing shortages (e.g., Schimtt et al., 2022), and many teachers felt increased job burnout and stress, in part because they believed the pandemic forced them to work more hours (Diiberti et al., 2021). These conditions have serious implications for the ability of schools to support research for two reasons. First, there are simply diminished staff resources to focus on research. Second, even among remaining staff, research is likely to be deprioritized to make room for other more immediate challenges (such as assuring compliance with local health mandates, ensuring that school lunches were distributed). Additionally, staff turnover has potential implications for human capital and institutional knowledge; those individuals who were familiar with the particulars of study planning or implementation may no longer be working in a particular school or district.

### *Limited Availability of Good-Quality Administrative Data*

In addition to the primary source data described above, education research studies regularly make use of a variety of administrative data and standard measures routinely collected by



schools and districts, including assessments, course grades, attendance, student surveys, among many others. The pandemic limited researchers and educators’ ability to collect high-quality data on many, if not all, of the measures central to education research. Beyond issues with the quality of assessment data (see below), a nationally representative survey conducted by the Center for Reinventing Public Education (CRPE) found that only half of districts monitored engagement through attendance tracking or one-on-one check-ins, and only around 40 percent of districts set expectations that teachers collect student work, grade it, and include it in final course grades (Gross & Opalka, 2020).

## Methodological Challenges Posed by Large-Scale School Disruptions

The framework developed by Campbell and colleagues (Campbell & Stanley, 1963; Cook & Campbell, 1979; Shadish et al., 2002) indexes 37 distinct threats to internal validity, construct validity, statistical conclusion validity, and external validity. Drawing from this framework, we focus on a subset of 10 validity threats that we believe are particularly salient during large-scale school disruptions such as the COVID-19 pandemic (see Table 1.1). Below, we briefly describe the 10 validity threats and explain how COVID-19 era schooling conditions might have heightened these threats in the studies conducted under these conditions.

**Table 1.1. Categories of Validity Threats**

Type	Definition	Specific Threats
Internal Validity	The extent to which the estimated effects of a program or intervention correspond to true causal effects	<ul style="list-style-type: none"> <li>• Selection bias</li> <li>• Attrition bias</li> <li>• Instrumentation bias</li> <li>• History bias</li> </ul>
Construct Validity	The extent to which the enacted treatments and measurement constructs faithfully represent the intended treatments and measurement constructs (Gorin 2006)	<ul style="list-style-type: none"> <li>• Inadequate explication of constructs</li> <li>• Treatment diffusion</li> </ul>
Statistical Conclusion Validity	The extent to which statistical methods are used appropriately	<ul style="list-style-type: none"> <li>• Low statistical power</li> <li>• Unreliability of treatment implementation</li> </ul>
External Validity	The extent to which inferences about effectiveness can be generalized to other individuals, settings, instruments, or program variations, (Briggs 2008)	<ul style="list-style-type: none"> <li>• Interaction of the causal relationship with settings</li> <li>• Interaction of the causal relationship over treatment variations</li> </ul>

SOURCE: Shadish et al, 2002

### *Threats to Internal Validity*

Internal validity describes the extent to which the estimated effects of a program or intervention correspond to true causal effects (Shadish et al., 2002). There are four primary threats to internal validity that are particularly relevant for our analysis. Selection bias is likely to arise if there are systematic differences between those who participate in an intervention and those who do not. This will occur, for example, if schools or districts that have more resources, more stable or resilient infrastructure, lower exposure (e.g., lower COVID transmission rates) or greater organizational capacity consent to participate in a study and implement an intervention. *Attrition bias* is likely to arise because of shifts in enrollment and staffing. Large numbers of students transferred out of public schools during and after pandemic-induced school closures (Dee & Murphy, 2021). Beyond changes in enrollments, student and staff absenteeism increased and schools reported trouble adequately staffing their positions (IES, 2022b). Dropout and absenteeism can lead to attrition bias if students or staff who remain in schools may be systematically different from those who withdrew. *Instrumentation bias* is likely associated with disruptions to state assessment systems. There were drastic changes to states' summative assessment programs in 2019–2020 and 2020–2021. Spring 2020 summative assessments were suspended in all states, and spring 2021 assessments were administered in a variety of modes (e.g., remotely, in-person). Some states administered abbreviated summative assessments while others used benchmark assessments like MAP Growth (Bruno & Goldhaber, 2021). These changes in instrumentation may be conflated with an exposure effect (internal validity). *History bias* will be a particularly salient threat to internal validity if an intervention or policy was enacted over a period that aligned with the pandemic. In such a situation, it is difficult—if not impossible—to disentangle the effects of the intervention on student outcomes from the confounding effects of the pandemic itself.

### *Threats to Construct Validity*

Construct validity describes the extent to which the enacted treatments and measurement constructs faithfully represent the intended treatments and measurement constructs (Gorin, 2006; Shadish et al., 2002). There are two primary threats to construct validity that are particularly relevant for our analysis. *Inadequate explication of constructs* may arise because of the disruptions to assessment programs described above. When assessments are altered or adapted, there is the possibility that these revised assessments do not capture the same aspects of the construct as the intended measures. *Treatment diffusion* may be induced by the drastic and sudden changes in school contexts and procedures that occur during unplanned and prolonged disruptions to schooling. As one specific example, the federal government allocated billions of dollars to state education agencies through the Elementary and Secondary School Emergency Relief (ESSER) Fund with the goal of supporting school restart and recovery. This money was allocated to programs (including high dosage tutoring programs, see Robinson et al., 2021) that,

in some cases may alter or overlap with the intervention activities that comprise the treatment or control conditions.

### *Threats to Statistical Conclusion Validity*

Statistical conclusion validity describes the extent to which statistical methods are used appropriately (Shadish et al., 2002). There are two primary threats to statistical conclusion validity that are particularly relevant for our analysis. *Low statistical power* may be induced by the shifts in enrollment and staffing described above. Heightened attrition can reduce statistical power. While attrition bias may arise if there is differential attrition, even if there are not systematic differences in study withdrawal or data availability across study conditions (e.g., if those that are participating in an intervention are more likely to have missing assessment scores or to withdraw from school entirely), widespread missing data or heightened attrition reduces the statistical power of a study to detect a relationship between an intervention or policy and an outcome. *Unreliability of treatment implementation* may arise because of how schools and systems altered interventions in response to the pandemic. For example, interventions designed for conventional classroom settings were frequently reconfigured to be implemented online. Furthermore, because schools reopened for in-person instruction on different timelines, implementation may have varied from setting to setting.

### *Threats to External Validity*

External validity describes the extent to which inferences about effectiveness can be generalized to other individuals, settings, instruments, or program variations (Briggs, 2008; Shadish et al., 2002). In the international development literature, Nene Odjidja and Alves do Reis (2021) note that because humanitarian settings are so dynamic and idiosyncratic, the generalizability of findings from research conducted in such settings may be limited and that findings may not replicate in different geographic areas or among different populations. There are two primary threats to external validity that are particularly relevant for our analysis. *Interaction of the causal relationship with settings* may occur because effects found in one site may not hold if the study were conducted in another site or at another point in time (including timepoints more temporally distant from the height of the pandemic). As mentioned above, because schools reopened for in-person instruction on different timelines, implementation of interventions may have varied from setting to setting given differences in instructional modes. Communities and schools were also differentially vulnerable to pandemic effects based on a combination of ecological, social, health and economic factors (Nayak et al. 2020). This raises questions about the extent to which effects identified by studies conducted in this era could be expected to hold in other settings or with other program instantiations. *Interaction of the causal relationship over treatment variations* may occur because effects found with one specific variation of an intervention or policy might not hold with other variations of that intervention or

policy. Many schools altered interventions in response to COVID-19, inducing some of the idiosyncrasy in program implementation noted by Nene Odjidja and Alves do Reis (2021).

## The Current Study

Throughout the pandemic, researchers had to think creatively about how to approach the design and implementation of education studies, often having to make adaptations and design pivots in response to practical and methodological challenges. Because these pivots were made reactively in the context of a worldwide crisis, there was little opportunity to think collaboratively about the best ways to ensure study feasibility and to mitigate the threats to validity.

This study is a first attempt to describe common challenges and successes with the objective of improving decision making by education researchers during future periods of disruption. We summarize themes from interviews with researchers who were fielding or planning IES-funded Exploration, Development and Innovation, Initial Effectiveness and Efficacy and Replication<sup>2</sup> studies during the pandemic. We find that researchers made pivots to address practical challenges and protect the feasibility of their studies. We also find that researchers took precautions, where possible, to understand and bolster internal validity. However, while the adaptations researchers made to their studies in response to practical and methodological challenges mitigated some threats to validity, the adaptations they chose frequently surfaced additional validity threats—particularly to construct validity and external validity.

The remainder of this article is structured as follows: First, we describe our data and methods. Then, we summarize our findings, including the practical challenges faced by researchers including issues with intervention feasibility, difficulty with recruitment, issues with data availability and data quality, and concerns with extrapolation and generalization. In the next section, we discuss these results, connecting our findings back to the practical and methodological challenges described in the international development literature. We conclude with recommendations for future studies conducted in times of prolonged unplanned school closures or other large-scale disruptions.

---

<sup>2</sup> Though IES has altered the goal structure several times since 2017, we use the goals and project types that appeared in the project summaries on the IES website throughout this analysis. Exploration grants, as defined by IES, identify malleable factors associated with student outcomes. Development and Innovation grants seek to develop interventions and improve existing interventions. Initial Efficacy studies include those that 1) examine interventions that have not been rigorously evaluated previously, 2) test the longer-term impact of interventions shown to have beneficial impacts on student outcomes, and those that 3) rely on retrospective data to test the impacts of past interventions. Efficacy and Replication studies examine the extent to which interventions found to have beneficial education outcomes on student outcomes replicate in other contexts and under different study conditions. A fifth goal, Measurement, is not included in the current study.

## Chapter 2. Data and Methods

---

In this qualitative study, we draw upon interview data to understand how researchers responded to turbulent environments and why they made the decisions they did. We purposively elected to focus on research studies funded through IES’s National Center for Education Research (NCER), both because of the breadth of research sponsored by NCER and because the details of funded studies are publicly available. We intentionally included studies on a wide range of topics, grade levels, and subject areas (e.g., math, reading, history, science).

We used a three-step process to identify interview participants. First, using publicly available information on IES-funded Exploration, Development and Innovation, Initial Efficacy, and Efficacy and Replication studies in fiscal years 2017 through 2021, we created a sampling frame from which to sample studies. Relevant studies for potential inclusion were identified through searches of the IES website, which contains project summaries for all funded projects. Available information includes principal investigator, grant type, award period, research design, and key measures. We excluded studies that were scheduled to conclude (meaning outcome data had already been collected) before the initial pandemic school closures in March 2020 as well as studies that did not rely on school or system administrative data (including state assessment data) for at least one study outcome. In total, we identified 85 studies that met our initial inclusion criteria.

**Table 2.1. Categories of Cases for Sampling**

<b>Case Category</b>	<b>Description</b>
Category 1	Studies that began prior to Spring 2020 school closures and were designed to include only one outcome data collection opportunity (n=3)
Category 2	Studies that were planned prior to Spring 2020 school closures but implemented in the Fall after Spring 2020 school closures (n=3)
Category 3	Studies that began prior to Spring 2020 school closures and were designed to include both pre-pandemic and post-school closure outcomes (n=3)
Category 4	Studies that were planned and implemented after Spring 2020 school closures (n=3)
Category 5	Studies that began after schools reopened in Spring 2021 (n=3)

Second, informed by the common study features outlined in Hedges and Tipton (2020), we conceptualized studies as generally including five research stages, include recruitment, random assignment, intervention or policy implementation, implementation data collection, and outcome data collection. We developed five categories that were characterized by the interaction between the research stage and the time point at which the pandemic began to influence them (Table 2.1).

These categorizations primarily reflect different starting points in time (rather than presenting distinct individual or study characteristics).

Category 1 includes single cohort studies that were influenced during outcome data collection phase. Category 2 includes studies that had been planned (and perhaps study sites had been recruited) but had not yet gotten to the intervention implementation stage until after spring 2020 school closures. Category 3 includes multi-cohort studies and longitudinal studies where outcome data collection was intended to happen both pre-COVID and after COVID-19 school closures. Category 4 includes studies that had recruitment, randomization, implementation, and outcome data collection happen after initial school closures, but had some part of their implementation stage happen when students in participating schools were learning remotely. Finally, Category 5 includes studies that were conducted after the most acute impacts of the pandemic subsided and schools were largely re-opened for in-person learning.

Third, we mapped each of the included studies from our sampling frame to one of these five categories. We reached out to leads of randomly selected studies within each category, until we reached three studies in each category. In total, we reached out to 18 study leaders and 15 agreed to participate. Our sample size falls within Hennink and Kaiser’s (2022) findings regarding adequate sample size in qualitative research, which suggests that 9-17 interviews are sufficient to reach saturation within relatively homogenous populations and narrowly defined objectives.

**Table 2.2. Descriptions of Included Studies**

Type	Category and Count
Goal	<ul style="list-style-type: none"> <li>• Exploration (n=2)</li> <li>• Development and Innovation (n=4)</li> <li>• Initial Efficacy (n=4)</li> <li>• Efficacy and Replication (n=5)</li> </ul>
Student or Teacher Focus	<ul style="list-style-type: none"> <li>• Student outcomes (n=7)</li> <li>• Teacher development (n=8)</li> </ul>
Grade Level	<ul style="list-style-type: none"> <li>• Elementary grades (K-5) (n=8)</li> <li>• Middle grades (6-8) (n=4)</li> <li>• High school grades (9-12) (n=3)</li> </ul>
Urbanicity	<ul style="list-style-type: none"> <li>• Urban (n=7)</li> <li>• Rural (n=3)</li> <li>• Multiple settings (n=5)</li> </ul>
Race/Ethnicity	<ul style="list-style-type: none"> <li>• Three studies mention a focus on a racially diverse sample</li> </ul>
Poverty	<ul style="list-style-type: none"> <li>• Four studies mention serving students experiencing poverty</li> </ul>

Table 2.2 highlights some descriptive information about the research participants. The majority of the studies included were Efficacy studies, and all but the two Exploration studies

included a planned experimental or quasi-experimental design. About half of the included studies examined the efficacy of interventions designed to improve student outcomes, and the other half were designed to improve teaching through coaching and professional development. The included studies also involved a wide range of student populations, including students in rural suburban and urban areas, and students from across the K–12 grade span. States from all geographic regions (Northwest, Southwest, Northeast, Midwest, Southeast) were represented, and two studies were multi-state studies. Many studies did not report demographic information on their participants, but those that did often reported samples as racially, ethnically, and economically diverse.

In spring and summer 2022, we conducted in-depth, semi-structured interviews with principal investigators (n=15, with three studies in each of the five categories, as noted above) and transcribed audio recordings prior to coding and analysis. Our interview protocol asked respondents about their study design and context, changes to school conditions, recruitment and sampling strategies, research design pivots, use of outcomes data, comparison of pre- and post-COVID-era outcomes, generalizability, and general successes and challenges in conducting research in turbulent settings. We began by coding interview data across different levels of abstraction, using Dedoose qualitative research software. For example, we coded for elements of study context (study intent, school level, study design, timeline), changes in instructional context (changes in conditions, modification of metrics and data, instructional model, staff turnover, fidelity of implementation, district priorities, use of funds), recruitment and sampling (sampling plan, recruitment efforts, issues, attrition), design pivots (research questions, hypotheses, counterfactual conceptualization, design and analysis plan), outcomes data (planned data sources, administration issues, data modifications), pre/post pandemic comparison (plans to compare or combine data, issues, modifications), and extrapolation and generalization (to pre-pandemic, to future). We then analyzed coded data using cross-case meta-matrices (Bush-Mecenas & Marsh, 2018; Miles et al., 2020) to understand prevalence and patterns among these quotes, paired with thematic analysis (Braun & Clarke, 2012) along the categories of validity (see Table 1, Shadish et al., 2002). To enhance the internal validity and accuracy of findings, we compared interview data with existing grant documentation. Specifically, we integrated grant proposal documentation to document study design and compare to described planned design and activities.

## Limitations

There were several limitations to our data collection and analyses. We focused our study on IES-funded research that relied on state or district administrative data for key measures of study outcomes. Because of this, most studies included in this analysis are Initial Efficacy or Efficacy and Replication studies. These studies are not meant to be representative of all IES studies or studies funded by other agencies. Relatedly, while we randomly selected studies for interviews,

there was some nonresponse, and we cannot be sure that those we did interview are representative of all studies funded by IES that were under the Exploration, Develop and Innovation, Initial Efficacy, and Efficacy and Replication agency goals. Third, we were unable to collect additional information on study implementation from school or district partners. While such data would have provided a richer understanding of how research programs were altered during the pandemic, the researchers we interviewed were reluctant to grant permission for such interviews, given concerns about capacity, burden, and sustaining research partnerships. Finally, because many of these research programs are actively funded by IES, confidentiality and anonymity were assured to all research participants. This included mitigating the possibility that the included studies could be identified by inference. As such, we are limited in the amount of study information and context we can provide. Although this study is best understood as exploratory, it is an important first attempt to describe common challenges and successes with the objective of improving decision making during future periods of disruption.



## Chapter 3. Findings

---

In this section, we provide our findings. Although our interview design and sampling approach was guided by the assumption that researchers' responses to the COVID-19 pandemic would vary based on research stage and the time point at which the pandemic began to influence them, our analyses instead surfaced patterns based on practical and methodological challenges that surfaced for researchers regardless of where they were in study planning and implementation. Additionally, while the existing literature in international development describes practical and methodological challenges separately, the researchers we interviewed for this study frequently described them as interconnected. Specifically, the researchers we interviewed framed the issues they faced as practical challenges with methodological implications. For this reason, we organize our findings around the three practical challenges described by researchers: (1) issues with intervention feasibility caused by situational complexity, (2) difficulty with study recruitment, and (3) issues with data availability and concerns about data quality. In the context of these challenges, we summarize common themes that emerged from researchers' pivots and adaptations. We then describe the associated methodological challenges that arose, framing those findings in terms of potential threats posed to validity. Finally, we conclude our findings section with a discussion of researchers' reflections on extrapolation and generalization, and the extent to which they believed that research conducted in the context of the pandemic was likely to be comparable to previously conducted studies, or to studies conducted several years from now, when the immediate impacts of the pandemic have subsided.

### Challenges To Intervention Feasibility Caused by Situational Complexity

#### *Threats to Intervention Feasibility Were Largely Caused by Instability in the Institutional Environment*

About half of our respondents indicated that instability in the institutional environment (including instructional resources, staffing, working conditions, technology, structure of the school day, and the school building) during the pandemic decreased the feasibility of their intended interventions. Most obviously, school closures as well as virtual and hybrid instruction interrupted or modified interventions by altering the structure of the school day and the physical spaces in which teaching and learning occur (especially for studies using traditional models of school-based classroom interaction as a variable of interest). One researcher referred to these aspects of the institutional environment as the underlying organizational infrastructure of schools, and noted,

The infrastructure in schools has been stressed...I think we've learned a lot more about ... how fragile that infrastructure is and how unchangeable and inflexible it is. And that's true of high schools, I think elementary and middles can flex easier than high schools. Their structure is years old and rarely has looked any different.

Most respondents shared similar observations about the full, system-level disruptions brought on by the pandemic: teacher absenteeism and retirements, lack of substitute teachers, shifting daily work for administrators to cover classrooms, increased student absenteeism, and students' heightened need for support all were identified as factors affecting the feasibility of ongoing research. One researcher noted the pandemic led to widespread "instability" due to "constant shifts" in staffing. This researcher continued,

For me the bottom line is we need to learn how to do applied research within these kinds of settings. As a researcher, I really wish it wasn't like this. It's really hard to control for all the variables that come into play, or even know how to... I think the big question for me when we meet as a team is, "What else should we be measuring that we're not measuring?" because we're not collecting data informally or formally to be able to understand and explain the context.

### *Researchers attempted to Mitigate Feasibility Challenges through Flexibility in Research Design*

In response to feasibility challenges, researchers took a variety of approaches, often depending upon the intervention itself and the phase of research. Almost all respondents reported that the researchers need to modify study timelines. Over half of the studies in our sample were delayed, with two studies ending prematurely due to inability to implement or measure the intervention under virtual instruction. Other researchers, especially those with studies scheduled to begin later in the pandemic, opted for a lengthened development period and delayed implementation. For three respondents, pilot studies and implementation analyses were conducted while awaiting suitable conditions for experimental research.

Where possible, researchers also simplified or modified interventions "to be less burdensome." Modifications addressed limitations on teachers' time (e.g., shorter professional learning segments), increased student need (e.g., a greater proportion of students needing tiered academic or behavioral interventions), and schools' use of virtual instruction (e.g., modifying instruction for use in virtual settings). In one example, a researcher described moving their intervention to an online platform and how this benefitted the implementation of the intervention, research data collection, and future research. As this researcher explained,

Making the materials accessible via the internet was very successful and I think it can be long-lasting. ... You can update your materials very easily when you're dealing with an online environment. You can get to students who may be absent more easily because you get the response right away and so you can follow up with the teachers to work on your response rates that way. I think that moving over to an online environment for the intervention and the data collection was something really positive that happened.

On the whole, almost all the researchers in our sample described flexibility as a strength in addressing shifting contexts. In the words of one researcher,

It is important to be flexible with the folks that were working on projects, to recognize ... that we're all wrestling with challenges.

### *Institutional Instability and Researcher Responses to Feasibility Challenges Posed Particular Threats to Internal, Construct, and External Validity*

Overall, the instability in the institutional environment identified by study participants poses threats to construct, internal, and external validity. As regards internal validity, the shifts to the basic structures of schooling present threats of history bias as it is difficult, if not impossible, now to disentangle the effects of the intervention from those of the pandemic. Turbulence in school settings also has the potential to affect the treatment or control condition, presenting the possible threat of treatment diffusion. Researchers who modified study timelines by delaying recruitment or implementation took steps to mitigate these validity threats by waiting for more “normalized” conditions to diminish the threat of history bias and implement an intervention or a policy with more fidelity to the intended model.

Researchers that simplified or modified interventions, on the other hand, may have solved some practical issues with launching and conducting a study at the cost of raising some additional threats to construct validity and external validity. Treatment adaptations present challenges to construct validity because the effects of the enacted treatment might not reflect the effects of the intended treatment. Additionally, there is the possibility of an interaction of the causal relationship over treatment variations, such that an effect found with the treatment as modified for the pandemic context might not hold with other variations of that treatment (e.g., Shadish et al., 2002), posing a threat to external validity.

## Difficulty With Study Recruitment

### *Difficulties with Study Recruitment Were Largely Caused by Limited School and District Resources as Well as Shifting Priorities*

Researchers experienced serious challenges to recruitment, exacerbated by limited district and school staff time as well as competing priorities that often superseded research participation. In about a quarter of the studies in our sample, recruitment challenges manifested as districts or schools dropping out entirely or refusing to enforce randomization. At the time of our interviews, many of the studies were actively recruiting for the experimental portions of their research and continuing to find research access and recruitment very challenging. Three respondents reported that, while districts were generally supportive in concept, they limited researchers from directly contacting schools or teachers, leading to substantial recruitment challenges. Another researcher noted that the influx of federal stimulus funds may have made “free” interventions accompanied

by research, which are often offered as part of evaluations, less attractive, which also presented recruitment challenges.

Situations arose where the district, or sometimes school leaders, sought to protect teachers from increased demands on their time amid widespread teacher burnout. One researcher identified the scope of excessive burden on teacher time as follows:

School-based personnel are so overwhelmed with not only trying to do their own jobs but trying to do their own jobs plus the 50 percent of the job that somebody else was doing that left. ... So they're doing really a tremendous number of things just to keep the school afloat. ... they also have less time than ever to dedicate to getting them [interventions] to work. ... That's like balancing the burden and knowing, even for research questions that the schools still see as important, how to carve out that time seems to be a real challenge.

Substitute teacher shortages further constrained teachers' participation in research. One researcher summarized this tension: "I think the biggest thing is schools are so overwhelmed right now that I think that the hardest thing is just to do research in schools is to be really respectful of where the schools are at."

### *Researchers Addressed Recruitment Challenges by Focusing on Partnerships and Allocating Funding to Support Staffing and Incentives*

To address these challenges, researchers took the approach of developing stronger partnerships within districts to build support for research and participation. First, most of the researchers in our sample discussed recognizing the main competing priorities facing districts. One researcher described this approach as, "realizing that research is not at the top of the agenda of districts and schools who are trying to put out fires." Five researchers reported offering or increasing incentives for participation. Noting that administrators often had less time to advocate for research, three researchers said they created new liaison positions to help with recruitment. In these studies, researchers hired retired school staff as research liaisons or teacher recruiters. These individuals were knowledgeable about the district and school context and priorities and better able to communicate the importance of the research to local staff. Two respondents stated that the nature of researchers' relationships with districts reportedly shifted. As one researcher described their pivot:

I think we really pivoted to try to do more like a Research-Practice-Partnership approach in terms of asking the schools, "Well what do you want? What do you need? What would be helpful for you?" And I think that was well received in general and I hope that that resulted in, maybe not us collecting research data, but in the school being able to utilize some of the resources that have been created.

These conversations sometimes led researchers to supplement their planned studies with pilot studies or dosage studies, or implementation and lessons learned analyses.

## *Recruitment Difficulty and Researcher Responses to Recruitment Challenges Posed Particular Threats to Internal, External, and Statistical Conclusion Validity*

Overall, the instability in the institutional environment identified by study participants poses threats to internal, external, and statistical conclusion validity.

In particular, schools and systems that have the capacity to support research may differ systematically from those who do not. This has two different implications for validity. First, in multi-site studies (those conducted in multiple schools or districts), if a subset of sites opts not to participate because of a lack of capacity, this potentially leads to attrition bias or selection bias. Second, if a planned or intended site is unable to participate and a new site is recruited, there are implications for external validity because the characteristics of those systems who do participate might restrict the populations to which results can be generalized.

Another threat to validity was explicitly identified by study participants. Over half of the researchers we spoke with noted that recruitment challenges were likely to result in smaller realized study samples. This potentially poses a threat to statistical conclusion validity by making the study insufficiently powered to detect effects.

## Data Availability and Concerns About Data Quality

### *Difficulties with Data Availability and Concerns about Data Quality Were Largely Caused by Changes to State Assessment Programs*

A common challenge was the lack of data availability during the pandemic. In March 2020, state testing holds meant that no statewide assessment data were available, which presented a substantial challenge to studies that were designed to make use of (if not solely rely upon) these data. However, a second challenge commonly described by the researchers we interviewed was low or uneven participation rates in assessments even after testing holds were lifted in spring 2021. One researcher described the challenge of limited student participation in state testing, stating, “If attrition [from our study] means that they [students in our sample] didn’t take the state tests, that has nothing to do with the actual study; that has everything to do with COVID.”

Another researcher emphasized the importance of interrogating the complexity of noise in state testing data. As this researcher shared,

It was pure chaos, but we still got impact results even amongst that chaos so in the grand scheme of things, I think it was a successful implementation, we’re just trying to figure out how to make sense of the noise. It wasn’t random noise, it was definitely COVID noise.

Challenges arose with use of other forms of data as well. About a quarter of researchers noted that virtual instruction precluded planned classroom observation data crucial to studies of instructional practice and classroom interaction. Virtual observations, in these studies, were inadequate measures of interaction and practice, given the substantive differences to instruction

in virtual versus in-person settings. Even administrative data, like attendance, were sometimes altered by districts in response to the pandemic context. For example, one researcher noted they needed to request more fine-grain data than usual, given shifts in metrics. As this researcher explained,

Everybody measured it [attendance] differently. And even now, in reality, there's differences. You have to be really careful when you get that data and know exactly what they count as a day and not a day ... At this point, we don't get summarized data; we ask for the individual data on the individual kids so that we can do what we want to do with it.

### *Researchers Addressed Challenges in Data Quality and Availability by Modifying Their Outcome Measures*

In response to data quality challenges, researchers took two main approaches. About a quarter of respondents reported altering the key outcome measures of the study, shifting away from a reliance on statewide summative assessments and placing more emphasis on other assessments such as the PSAT. Three researchers noted they built their own assessments or surveys, making use of internet and smartphone-based technology to facilitate high response rates. Researchers who pivoted towards purpose-built assessments and surveys administered online, believed these data better reflected the aims of the intervention. Other forms of data collection included recorded classroom instruction (where school access was limited), as well as qualitative data collection to better understand implementation.

### *Data Quality and Researcher Responses to Data Quality Challenges Posed Particular Threats to Internal, Construct, and External Validity*

Overall, the instability in the institutional environment identified by study participants poses threats to construct, internal, and external validity. Many of these threats were explicitly identified by study participants. About a third of researchers raised concerns about the implications of missing data for the internal validity of these state test results, with a particular focus on concerns of attrition bias and selection bias. For example, one researcher noted patterns in which students did not complete the state test due to parent opt-out. In this study, the researcher noted that students who had selected in-person instruction were more likely to take the state test than their peers who opted to continue virtual instruction. These two populations of students, however, differed systematically both on prior performance and in race/ethnicity. As one researcher summarized, “the problem is that through our analysis of state results is like, that missing data isn't like random.” Another researcher raised concerns about the use of state tests for research given possible attrition bias, as students who were tested may have differed systematically from those who did not take state tests, and instrumentation bias, as students might be tested virtually versus in-person based on whether they were accessing virtual or in

person instruction. As this researcher shared, “We’re concerned not just, ‘Will the state tests be there?’ but ‘how well we can trust those results?’”

Modifying outcome measures is one way to mitigate the threats to validity posed by concerns that missing state testing data is likely to induce attrition bias and selection bias. However, modifying outcome measures also raises some threats to validity that merit consideration. First, there is the threat of instrumentation bias that arise any time an assessment program is changed, because changes in assessment may be conflated with a treatment exposure effect. Second, it is unclear to what extent new assessments might vary from the intended constructs for the study, potentially raising questions about construct validity. Finally, using new assessments may have consequences for external validity, because using study-specific measures may make it difficult to compare effects across studies, or impede future researchers’ capacity to replicate study results or include such studies in meta-analyses.

## Concerns with Extrapolation and Generalization

The challenges described above have immediate implications for the quality and rigor of individual studies. While researchers were able to adjust research plans to sustain individual studies, another concern is external validity, or the extent to which research findings can be generalized across studies or to other settings.

About half of the researchers in our sample expressed concerns that research studies from the pandemic period could not be generalized to other contexts, due to the pandemic’s impacts on school learning environments, staffing, and differential impacts on specific student populations. As one researcher noted, “The learning environment was so incredibly different than anything we’ve seen in the past and hopefully anything we see in the future”—referring not only to the shift in conditions for the treatment condition, but also to the counterfactual. Indeed, understanding the counterfactual itself was difficult. Three researchers noted that district responses to the pandemic (e.g., shutdowns, virtual, hybrid, in-person) amplified existing inequities. In such settings, it is difficult to isolate the effect of the intervention, where the counterfactual itself may have shifted in varied and inconsistent ways. While thoughtful about the shifts in the counterfactual, no researchers in our study described systematically measuring the business-as-usual conditions under the pandemic.

Systematic analysis of how hybrid or virtual settings might have influenced peer effects were uncommon, despite awareness of such potential issues among researchers. Nonetheless, six researchers suggested that changes like virtual learning and limited school staffing that would likely persist in the future, enhancing generalizability of pandemic research. One such researcher shared,

In some ways the information during the pandemic is unique and those conditions aren’t exactly the same, but some of those conditions are going to continue to be part of schooling and should be addressed.

Four researchers believed that the pandemic conditions merely narrowed the range of generalizability, such that findings could only be generalized to similar settings, grade levels, or types of interventions.



## Chapter 4. Discussion and Summary

---

During the pandemic, researchers struggled to strike a balance between the research studies that were intended and those that could realistically be accomplished. Once the COVID-19 pandemic upended educational systems, there were effects on program and policy implementation and how research activities were enacted. In some cases, large-scale disruptions like the pandemic even made studies entirely infeasible.

Our research shows many parallels between school-based studies conducted during the pandemic and the literature on practical and methodological challenges of conducting research during humanitarian crises and conflict. Specifically, the researchers we interviewed framed the issues they faced as practical challenges with methodological implications. Overall, researchers we spoke to were concerned with sustaining their studies in some form throughout the pandemic, and often evaluated their adaptations in terms of whether particular pivots would allow them to proceed with a research agenda which might otherwise be impossible. Specifically, researchers were primarily concerned with addressing challenges to intervention feasibility caused by situational complexity, difficulty with study recruitment, and challenges with data availability and concerns about data quality. The practical challenges mentioned by the researchers we interviewed are generally consistent with the practical challenges outlined in international development literature (e.g., Bakrania et al., 2021; Puri et al., 2017), though our respondents did not generally distinguish between challenges of accessing research populations and challenges presented by limited availability of good-quality administrative data.

There were also similarities in the methodological challenges faced by researchers planning or conducting education studies during COVID-19 and the international development literature, as regards threats to validity. Our findings suggest that education researchers were confronted with many of the same validity threats identified by Puri and colleagues (2017), and that researchers often made pivots and adaptations that addressed threats to the internal validity of their studies. However, our findings also suggest that these pivots and adaptations also had the unintended consequence of raising other threats to validity. For example, researchers addressed institutional instability in ways that mitigated threats of history bias but may have simultaneously exacerbated other threats to construct validity and external validity. Additionally, researchers addressed data quality issues by creating alternative assessments. This on one hand mitigated threats to internal validity posed by attrition bias or selection bias, but on the other hand raised the possibility that the enacted measurement constructs do not faithfully represent the intended measurement constructs.

Our findings make clear that concerns about generalizability and extrapolation were less of a priority for researchers during the pandemic. While some respondents acknowledged that the pandemic was an unusually turbulent time, there was little evidence that researchers gave much

systematic thought to how (and whether) the results of their studies are likely to be predictive of results in another specific school or district contexts (e.g., local effectiveness predictions, see Joyce and Cartwright 2020), whether the results of their studies should be synthesized with results of other studies, or the extent to which the results of their studies would ultimately contribute to advancing disciplinary knowledge more broadly.

In fact, commonly used strategies such as modifying or simplifying interventions, along with significant variation in local conditions, may also make generalizing findings to the post-pandemic period more challenging, and may complicate how educators use research results to inform practice. Large-scale disruptions like those caused by the pandemic suggest the features of schooling often assumed stable across studies and over time can be upended, making generalization within and across studies fraught. In some ways, addressing concerns about whether studies conducted during the height of the pandemic should and could generalize to the post-pandemic era may seem like a less immediate issue than practical issues around recruitment or data collection. However, if a study was focused on studying a policy or program that was designed to be delivered under standard operating conditions, understanding extrapolation and generalization to other periods or contexts is an important consideration, particularly given education policymaker's increasing interest in to better understanding which interventions promote positive educational outcomes, for whom, and under what conditions (e.g., IES 2022a). Even if research is motivated by improving assistance in response to a specific crisis, understanding the impacts of interventions and policies in other schools or districts (particularly those with different degrees of pandemic vulnerability or different policies on remote schooling), is important for informing local decisions about intervention or policy adoption.

## Conclusion

It is widely accepted that, even before the pandemic, there were myriad practical and methodological challenges that threatened both the viability of education research and the validity of research findings (e.g., Myers, 2014). As one researcher summarized, “logistics are the things that break almost every study.” However, there has never been a disruption to schooling in modern history as profoundly impactful and widespread as the COVID-19 pandemic. And while there is a growing body of work in international development about conducting research in fragile contexts, including in times of conflict or in the midst of humanitarian crises (Hassnain et al., 2021; Puri et al., 2017; Bakrania et al., 2021; Nene Odjidja & Alves do Reis, 2021), there is little research evidence on conducting studies of education policies and programs when faced with large-scale school disruptions.

This study begins to fill this gap by utilizing interview data to describe common challenges faced by researchers conducting studies during the COVID-19 pandemic, as well as the strategies employed by researchers to successfully navigate this turbulent time. Overall, this study finds the disruptions to schooling caused by the COVID-19 pandemic showed how difficult it can be for

researchers to maintain a balanced focus on potential threats to internal, construct, statistical conclusion, and external validity when conducting an education study under exceedingly turbulent conditions. In other words, the uncertain and unstable schooling conditions presented by the pandemic inevitably surfaced tensions among these aspects of high-quality studies, as researchers were forced to address issues reactively, without opportunities for collaboration or collective sensemaking (Grissom and Condon 2021).

Based on our findings and informed by emerging best practices in the international development literature (e.g., Nene Odjidja & Alves do Reis, 2021), we conclude by offering three suggestions to help researchers plan ahead for planning and conducting education studies in the context of future large-scale school disruptions. Considering these suggestions while planning mitigate both practical and methodological challenges posed by large-scale school disruptions. In particular, these suggestions encourage researchers to proactively consider issues of extrapolation and generalization, maximizing the potential for such studies to contribute to our broader understanding of how to improve teaching and learning.

**Flexible and adaptive designs are important for ensuring that research remains feasible and that findings remain useful.** Part of a flexible or adaptive design involves contingency planning, which may involve including a “cushion” in the study implementation schedule, or planned adaptations to data collection procedures if situations arise that compromise the feasibility of a particular procedure. Adaptive design also involves planning on using multiple measures or mixed methods for data collection. Several authors note that planning to use a mix of qualitative and quantitative methods can be a productive way to proactively respond to the practical and methodological challenges of conducting research during a large-scale disruption (Chirambwi, 2023; Nene Odjidja and Alves do Reis, 2021).

Flexible planning may also include allocating additional project resources to support recruitment, including financial resources for incentives, or products that are of immediate use to schools. This kind of planning may allay logistical issues, while ensuring research is ethically conducted.

**Establish and cultivate trusting partnerships with schools and districts.** Trusting partnerships ensure that researchers and educators agree with and are committed to the objectives of the study (López Turley and Stevens 2015). This can be critical in mitigating recruitment challenges and ensuring that research remains beneficial to participants as well as to the field. In particular, potential research participants may be more trusting of school system officials than outside researchers (Bartlett et al., 2017; Bruzzese et al., 2009), and some research suggests that school partnerships may increase the likelihood of successfully recruiting historically underserved students, including English Learners and students experiencing poverty (Alibali & Nathan, 2010; Bartlett et al., 2017; Lamb et al., 2001). Partnerships focused on recruitment can help to mitigate the possibility of selection bias. Additionally, strong partnerships can also mitigate other threats to validity by creating opportunities for effective communication between researchers, program implementers, and other district personnel (López Turley & Stevens, 2015)

so that researchers can better document and transparently report shifts in implementation and to understand counterfactual conditions more clearly.

**Document modifications to study designs, changes to instruments, and information about students' learning contexts.** Research conducted during the pandemic was conducted in the midst of systemic turbulence and instability. Our study found that generalizability and extrapolation were of significantly lower priority for researchers than practical issues of study feasibility, or mitigating threats to internal validity. However, documentation about these key variables is critical for understanding how these aspects of the study shape inferences about the effectiveness of a program or policy (Hill et al., 2022), and for increasing the credibility of local effectiveness predictions made by practitioners and policymakers (Joyce & Cartwright, 2020). In general, researchers suggest that such information is under-reported in academic papers (Hill et al., 2022), and though SEER standards do mention the importance of documenting implementation, there are not clear and consistent standards for publishing such contextual information across journals (Hill et al., 2022). Ultimately, this kind of documentation provides consumers crucial information to consider the external validity of the study and how this study may fit into a broader body of literature, including replication studies and meta-analyses. In addition, in documenting these shifts, we suggest that researchers intentionally and explicitly discuss the potential implications that in-stream adjustments to study design, implementation and analysis may have for internal validity, construct validity, statistical conclusion validity and external validity. While the complex ecosystem of schools and turbulent conditions will always necessitate making compromises, reporting these trade-offs will help to strengthen the field through more comprehensive knowledge in the presence of complex research conditions.

## Abbreviations

---

COVID-19	coronavirus disease 2019
CRPE	Center for Reinventing Public Education
ESSER	Elementary and Secondary School Emergency Relief
IES	Institute for Education Sciences
INEE	Inter-agency Network for Education in Emergencies
NCER	National Center for Education Research
PSAT	Practice SAT
SEER	Standards for Excellence in Education Research
WWC	What Works Clearinghouse
COVID-19	coronavirus disease 2019

## References

---

- Alibali, Martha W., and Mitchell J. Nathan. "Conducting research in schools: A practical guide." *Journal of Cognition and Development* 11, no. 4 (2010): 397-407.
- Bakrania, Shivit, Nikola Balvin, Silvio Daidone, and Jacobus De Hoop. "Impact evaluation in settings of fragility and humanitarian emergency." (2021). As of November 1, 2023: <https://www.unicef-irc.org/publications/pdf/Impact-Evaluation-in-Settings-of-Fragility-and-Humanitarian-Emergency.pdf>
- Bartlett, Robin, Tiffany Wright, Tia Olarinde, Tara Holmes, Emily R. Beamon, and Debra Wallace. "Schools as sites for recruiting participants and implementing research." *Journal of community health nursing* 34, no. 2 (2017): 80-88.
- Braun, Virginia, and Victoria Clarke. *Thematic analysis*. American Psychological Association, 2012.
- Briggs, Derek C. "Comments on Slavin: Synthesizing causal inferences." *Educational Researcher* 37, no. 1 (2008): 15-22.
- Bruno, Paul, and Dan Goldhaber. "What pandemic-related test waiver requests suggest about states' testing priorities." *Phi Delta Kappan* 103, no. 3 (2021): 48-53
- Bruzzese, Jean-Marie, Richard Gallagher, Sharon McCann-Doyle, Philip T. Reiss, and Neil A. Wijetunga. "Effective Methods to Improve Recruitment and Retention in School-Based Substance Use Prevention Studies." *Journal of School Health* 79, no. 9 (2009): 400-407.
- Bush-Mecenas, Susan, and Julie A. Marsh. "The DIVE approach: Using case-ordered meta-matrices and theory-based data displays to analyze multiple case study data." *Complementary research methods for educational leadership and policy studies* (2018): 33-56.
- Campbell, Donald T., and Julian C. Stanley. *Experimental and quasi-experimental designs for research*. Ravenio books, 2015.
- Chirambwi, Kudakwashe. "Rethinking research methods in protracted violent conflicts in Mozambique: fieldwork in complex emergencies." *Journal of Humanitarian Logistics and Supply Chain Management* (2023).
- Cook, Thomas D., Donald Thomas Campbell, and Arles Day. *Quasi-experimentation: Design & analysis issues for field settings*. Vol. 351. Boston: Houghton Mifflin, 1979.
- Dee, Thomas S., and Mark Murphy. "Patterns in the pandemic decline of public school enrollment." *Educational Researcher* 50, no. 8 (2021): 566-569.

- Diliberti, Melissa, Heather L. Schwartz, and David Matthew Grant. *Stress topped the reasons why public school teachers quit, even before COVID-19*. Santa Monica, CA: Rand, 2021. As of November 1, 2023: [https://www.rand.org/pubs/research\\_reports/RRA1121-2.html](https://www.rand.org/pubs/research_reports/RRA1121-2.html)
- Gorin, Joanna S. "Test design with cognition in mind." *Educational measurement: Issues and practice* 25, no. 4 (2006): 21-35.
- Grissom, Jason A., and Lara Condon. "Leading schools and districts in times of crisis." *Educational Researcher* 50, no. 5 (2021): 315-324.
- Gross, Betheny, and Alice Opalka. "Too Many Schools Leave Learning to Chance during the Pandemic." *Center on reinventing public education* (2020). As of November 1, 2023: <https://files.eric.ed.gov/fulltext/ED605576.pdf>
- Hur Hassnain, Lauren Kelly, and Simona Somma. "Evaluation in Contexts of Fragility, Conflict and Violence: Guidance from Global Evaluation Practitioners." (2021). As of November 1, 2023: <https://ideas-global.org/wp-content/uploads/2021/06/EvalFCV-Guide-web-A4-HR.pdf>
- Hedges, Larry, and Elizabeth Tipton. "Addressing the Challenges to Educational Research Posed by Covid-19." *Northwestern Institute for Policy Research Working Paper Series WP-20-47* (2020). As of November 1, 2023: <https://www.ipr.northwestern.edu/documents/working-papers/2020/wp-20-47rev.pdf>
- Hill, Heather, Kathleen Lynch, Joshua Polanin, Elizabeth Stuart, and Robert Ochsendorf. "Making Meta-Analyses Easy: Guidelines for Reporting on Classroom-Level Instructional Interventions." In *SREE 2022 Conference*. SREE, 2022. As of November 1, 2023: <https://sree.confex.com/sree/2022/meetingapp.cgi/Session/1624>
- Institute of Education Sciences (IES) (2022a). Standards for Excellence in Education Research. As of November 1, 2023: <https://ies.ed.gov/seer/>
- Institute of Education Sciences (IES) (2022b). School Pulse Panel. As of November 1, 2023: <https://ies.ed.gov/schoolsurvey/spp/>
- Inter-agency Network for Education in Emergencies (INEE). (2020). *20 Years of INEE: Achievements and Challenges in Education in Emergencies*. New York, NY. As of November 1, 2023: <https://inee.org/resources/20-years-of-inee>
- Jahan, Ferdous A., Nicole Zviedrite, Hongjiang Gao, Faruque Ahmed, and Amra Uzicanin. "Causes, characteristics, and patterns of prolonged unplanned school closures prior to the COVID-19 pandemic—United States, 2011–2019." *PLoS One* 17, no. 7 (2022): e0272088.

- Joyce, Kathryn E., and Nancy Cartwright. "Bridging the gap between research and practice: Predicting what will work locally." *American Educational Research Journal* 57, no. 3 (2020): 1045-1082.
- Lamb, Jacqueline, Kathryn R. Puskar, and Kathleen Tusaie-Mumford. "Adolescent research recruitment issues and strategies: application in a rural school setting." *Journal of pediatric nursing* 16, no. 1 (2001): 43-52.
- López Turley, Ruth N., and Carla Stevens. "Lessons from a school district–university research partnership: The Houston Education Research Consortium." *Educational Evaluation and Policy Analysis* 37, no. 1\_suppl (2015): 6S-15S.
- Miles, Huberman, and A. Michael Huberman. "Saldana.(2014). Qualitative data analysis: A methods sourcebook." *New York: Sage Publications, Inc* (2020).
- Myers, Edward L. "Taming Turbulence: An Examination of School Leadership Practice during Unstable Times." *Professional Educator* 38, no. 2 (2014): n2.
- Nayak, Aditi, Shabatun J. Islam, Anurag Mehta, Yi-An Ko, Shivani A. Patel, Abhinav Goyal, Samaah Sullivan et al. "Impact of social vulnerability on COVID-19 incidence and outcomes in the United States." *MedRxiv* (2020): 2020-04.
- Nene Odjidja, E. & Alves do Reis, A. (2021) Learning from adaptive evaluation approaches in dynamic conflict-afflicted and humanitarian settings. As of November 1, 2023: <https://idev.afdb.org/en/content/learning-adaptive-evaluation-approaches-dynamic-conflict-afflicted-and-humanitarian>
- Peele, Holly, and Maya Riser-Kositsky, "Map: Coronavirus and School Closures in 2019–2020," *Education Week*, last updated October 13, 2021.
- Plucker, Jonathan A., and Matthew C. Makel. "Replication is important for educational psychology: Recent developments and key issues." *Educational Psychologist* 56, no. 2 (2021): 90-100.
- Puri, Jyotsna, Anastasia Aladysheva, Vegard Iversen, Yashodhan Ghorpade, and Tilman Brück. "Can rigorous impact evaluations improve humanitarian assistance?." *Journal of Development Effectiveness* 9, no. 4 (2017): 519-542.
- Robinson, Carly D., Matthew A. Kraft, Susanna Loeb, and Beth E. Schueler. "Accelerating Student Learning with High-Dosage Tutoring. EdResearch for Recovery Design Principles Series." *EdResearch for Recovery Project* (2021). As of November 1, 2023: <https://files.eric.ed.gov/fulltext/ED613847.pdf>
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.



Slavin, Robert E. "Perspectives on evidence-based research in education—What works? Issues in synthesizing educational program evaluations." *Educational researcher* 37, no. 1 (2008): 5-14.

U.S. Department of Education. (2002a). *No Child Left Behind: A desk-top reference*. Washington, DC: Author. Available from <https://eric.ed.gov/?id=ED471334>