

July 2023

The Challenge of Execution: District Leader Insights on Evaluating Academic Programs, Products, and Services

Jon Fullerton
Brittany Miller

SUGGESTED CITATION

Fullerton J., and Miller, B. (2023) The Challenge of Execution: District Leader Insights on Evaluating Academic Programs, Products, and Services. Research Report. Cambridge, MA: Center for Education Policy Research, Harvard University

This research was supported by the Walton Family Foundation.

Table of Contents

| | |
|--|----|
| Summary Findings | 5 |
| Detailed Findings | 6 |
| Concluding Recommendations for District Engagement | 13 |
| Appendix A | 15 |

ABSTRACT

During the summer and fall of 2022, researchers at the Center for Education Policy Research (CEPR) at Harvard University conducted a series of interviews to explore district leaders' willingness to participate in evaluation efforts for academic products and services. There were three primary research objectives:

1. To determine what information district decision-makers need, want, and would find helpful in the identification and adoption of interventions and learning technology products and to better understand when in the procurement and decision-making process this information could be most useful.
2. To determine what barriers exist for schools and districts to pilot and test interventions before rolling them out system-wide, and which evaluation designs districts would be most willing to implement.
3. To understand what model of engagement is most desirable for districts to participate in rigorous evaluation work, including barriers to participation.

This study engaged 16 district leaders across the country in qualitative interviews. District leaders participating in this analysis represented 13 unique states with regional representation in the South (5), West (4), Southwest (4), Midwest (2), and New England (1). In total, leaders from participating districts serve 1.2 million students with individual district sizes ranging from about 25,000 to 200,000 students. All the leaders interviewed served a critical role in their district's decision-making processes for the procurement of academic products and services. Most participating leaders were either part of their district's leadership cabinet, or directly reported to a cabinet member. These interviews probed how district leaders currently incorporate evidence of effectiveness into their purchasing of academic products and services (i.e., academic interventions and core curricular materials), what evidence they wished they had, how willing they would be to have their district generate its own evidence through program evaluations, and what structures could be most helpful to support their decision-making moving forward.

Following the interviews, the research team also analyzed the studies that vendors provided to school districts during the procurement process. The purpose of this analysis was to better understand how vendors report on the effectiveness of their product, and to what degree these reports align with rigorous research methodologies.

Below, we summarize our findings and their implications for developing a network of districts focused on using high-quality evidence to select academic products and services. In Appendix A, we attach a brief outline of what such a network might look like and a timeline for its development.

Summary Findings

Two overarching findings stood out.

First, overcoming implementation challenges is seen by district leaders as a precursor to understanding the efficacy of an intervention, product, or service. The district leaders interviewed use multiple information sources to make decisions about academic products and services. However, due to the complexity of the environment in which they operate, the most important data to district leaders focuses on implementation fidelity and product satisfaction. In other words, districts ask, “Do teachers and schools use the program correctly?” and “Do they like it?” Because implementation fidelity is so difficult to accomplish, district leaders noted the challenge of getting to a point where evidence of efficacy could even be reliably generated or meaningful. For instance, if students are simply not showing up for a vendor-provided tutoring program, it is hard to evaluate whether the program would be effective if the students did show up. Furthermore, even if there was strong evidence of efficacy in other contexts, it would not matter for that district if they could not ensure strong implementation.

Second, while researchers typically want districts to try new interventions or products, districts would often benefit from doing less. Some of the district leaders noted that while their systems may use evidence to make decisions about what academic products and services to adopt, they rarely apply the same rigor to thinking about what to abandon. This leads to an accretion of programs over time that can overwhelm the system, perpetuating the lack of implementation fidelity.

The broader set of central themes from our interviews include:

- **Data Use:** District leaders we interviewed value the use of data when making decisions about how to invest resources to support student learning. However, the way they define data, the availability of reliable data, the timelines around data availability, and when data is needed to inform budget cycles all vary widely. For example, some districts valued data from vendors that showed alignment to student performance on state end-of-year summative assessments, whereas others described data analysis as primarily being focused on qualitative feedback from teachers after trying out the product they planned to adopt.
- **Implementation Fidelity:** The most common challenge district leaders reported in their ability to make system-level decisions about academic products was their ability to ensure strong implementation of the product or service. This need heavily influences the way district leaders make decisions about the purchase of new products or services, how they pilot and scale academic interventions and core curriculum, and when they may choose to abandon an existing product or service in their district.
- **Budget Constraints:** District leaders reported that, while they value evidence about the efficacy of academic products, budget constraints frequently constrain their ability to choose one product over another. This is further impacted by state and district policies that regulate how districts spend public funds to address student learning needs. District leaders also expressed concern regarding how their systems will make decisions as they approach the ESSER funding cliff in 2024, when federal grants will expire and districts will return to a context of more limited resources.
- **Leadership Matters:** One of the primary reasons named by district leaders for the adoption or abandonment of a certain academic product was changes in leadership, primarily with the superintendent. In these cases, the preferences of those higher in the hierarchy tended to outweigh any evidence of effectiveness that the district had generated—or, minimally, that evidence was not sufficiently visible to members of the leadership team to change direction.

- **Political Context:** District leaders frequently named the politics within their system as a primary driver when making decisions about academic products and services to address student learning needs. This is mainly due to the many stakeholder groups leaders need to engage in these decisions, further adding to the complexity.
- **Balancing Intervention and Core Instruction:** While the initial research questions for this study were focused only on academic interventions, district leaders expressed the importance of balancing a focus on decisions about academic interventions with a focus on decisions about core instruction.
- **Evaluation Networks:** All but one district leader interviewed expressed a desire to participate in an evaluation network in which researchers would help them—and other district leaders—incorporate evidence of effectiveness into the decisions they make about purchasing and implementing academic products to address student learning needs.
- **Feasibility of Testing Designs:** Even though district leaders may be interested in efficacy evaluation in theory, evaluations must be practical to execute within the district context. For an evaluation of an intervention to be meaningful, there must be a valid comparison group that did not receive the intervention. District leaders were asked which of 3 common approaches to identifying a valid comparison group for measuring efficacy were feasible in their districts. The 3 common approaches described were using a cut score to assign an intervention, using a lottery (randomized controlled trial), or using a phased implementation approach. Of the three approaches, the most feasible for district leaders was the phased implementation approach, followed closely by using a cut score. Only half of district leaders thought that the lottery, or randomized control trial, approach was feasible in their systems, with many leaders sharing concerns about equity with using this approach.
- **What to Abandon:** One of the most difficult questions for district leaders to answer was how they make decisions about what academic products and services to stop using, or to abandon, yet this is one of the primary needs district leaders reported. Several district leaders shared that focusing on what to abandon would be the most useful way to participate in an evaluation network, particularly given the current context: the upcoming funding cliff districts will face when ESSER funding expires.

These central themes speak to the complexity facing district leaders as they make decisions about the resources available to address student learning needs. We lay out more detail about each of these complexities below and discuss associated recommendations to guide the creation of evaluation networks that could provide support to district leaders in this work.

Detailed Findings

Data Use

District leaders value the use of data when making decisions about how to invest resources to support student learning. However, the way they define data, the availability of reliable data, and the timelines around data availability and budget cycles all vary widely.

District leaders repeatedly shared that they use data to inform procurement decisions for academic products. They also shared that they collect and analyze data during the pilot phase of implementing a new academic product to inform plans for scaling use of the product. When asked what kind of data they use to make these de-

cisions, the most common response was data received from the vendor. The research team analyzed the reports from vendors that our districts had received and found that, unsurprisingly, vendors only share positive findings. In some cases, the quality of the research is quite low. Examples of quality issues include very low N sizes, results that only show statistical significance in one grade level, or results that are focused on teacher or family satisfaction alone instead of any quantifiable student outcome results. Such findings might be characterized as marketing research instead of efficacy research.

District leaders also shared that they use third-party sources to decide whether to adopt a certain academic product, such as What Works Clearinghouse or Ed Reports. Yet in these cases, district leaders reported the limitations of what they were able to learn. For instance, one leader shared that Ed Reports focuses on alignment to standards as opposed to evidence of efficacy through rigorous research studies, while another leader shared that, “[What Works Clearinghouse] sometimes [has] a fairly limited pool of resources that have, you know, met the criteria....so that presents a bit of a challenge at times.”

Another barrier for district leaders being able to analyze meaningful student outcome data to understand the efficacy of a product or service is the availability of data. Districts often don’t have the data they need to evaluate products – particularly regarding implementation. Some of the primary barriers to data collection related to the difficulty of gathering necessary and consistent data from school sites to analyze centrally and the complexities associated with disentangling the effects of multiple, overlapping interventions or academic products that have been simultaneously implemented.

Finally, the timelines of when data are available to inform system-wide procurement decisions do not typically align with budget cycles. Because state-level summative assessment data is not available until the summer, but annual district budgets must typically be finalized early in the spring for the following school year, district leaders cannot typically leverage academic outcome data to inform whether to scale or continue use of an existing product or service.

Recommendations: Data Use

An evaluation network could provide support to district leaders in how to interpret results provided by vendors selling academic products, as well as what questions to ask vendors based on the “research” that is shared. District leaders are interested in studying the impact of academic products and services in their own district, particularly when they are able to learn alongside peer districts across the country that are using the same academic products or services. Finally, because the annual budget cycle does not align well with availability of state assessment data, district leaders expressed a desire to analyze and interpret data for products and services already in use in their system, enabling them to understand efficacy of products they currently use, as opposed to those they intend to adopt in the future. In any case, to be relevant, findings from an evaluation network need to be available for district leaders by early December to impact the following school year’s budget and planning cycle.

Implementation Fidelity

The most common challenge district leaders reported in their ability to make system-level decisions about academic products was being able to ensure strong implementation of the product or service. This need heavily influences the way our district leaders make decisions about the purchase of new products or services, how they pilot and scale academic interventions and core curriculum, and when they may choose to abandon use of an existing product or service in their district.

When district leaders described how they decided which academic product or service to purchase, the most common factor that weighed into their decision-making was how easily it could be implemented in their system. Some of the specific implementation needs leaders cited were how well their stakeholders (primarily teachers and school leaders) liked the product, the vendor’s willingness to work with district staff and make customiza-

tions based on district needs, and to what extent data from the product would be available and easily integrated into existing systems. In each of these cases, when probed about the relative importance of the efficacy of the product in decision-making, district leaders shared that this also mattered, but that if their system couldn't implement the product or service well, they would not be able to rely upon efficacy data.

Understanding this context within which district leaders operate colors many of the other findings and recommendations in this study. District leaders are not only grappling with whether there is reliable data regarding the efficacy of the product or service they are purchasing. They must also translate the conditions under which the product or service was efficacious elsewhere to the complex, political, and often unique environment in their districts.

As a result, district leaders typically responded that the purpose of pilots in their systems is not to understand the binary efficacy of a product, but instead to understand how to most effectively use the product or service within their context. Pilots of new products or services therefore typically focus on how to implement something new once the decision has already been made about what to purchase. As one leader shared, "I think we're mostly trying to learn about the ease of use to set up the structure ... I think we wanted to hear from teachers about how they were using [and] how they built the program... and what they were seeing as benefits." In many cases, district leaders refer to a "pilot" as a phased rollout designed to address implementation barriers prior to scaling—not a test of whether the product "works." Some district leaders described pilots as the comparison of two or more products or services to choose one to scale. Even in these cases, the focus of decision-making tends to be on which product is being implemented with the greatest ease in the system—not efficacy.

The focus on fidelity of implementation also feeds into the decision to abandon an academic product or service. This is most closely linked to instances when districts have purchased a relatively discrete educational technology service, such as artificial intelligence software or an on-demand tutoring service. Districts may find that teachers and students are either no longer using the product, or there was never uptake in the initial implementation. In these cases, the abandonment of the product is typically easy to make once the implementation data is known. Unfortunately, the money spent to scale a product used by few is lost forever.

Recommendations: Implementation Fidelity

While district leaders still did express a desire to understand the effectiveness of a product or service they were purchasing or had already purchased, the strong responses regarding the importance of working through implementation challenges must be considered. A district-driven research and evaluation network would support districts in understanding under what conditions the product or service shows a positive impact on student achievement and the usage and/or delivery patterns within their district. When implementation is not where it needs to be, the network would support districts in testing strategies to improve uptake and usage.

Budget Constraints

District leaders reported that while they value the evidence of academic products' efficacy, budget constraints frequently outweigh their ability to choose one product over another. This is further impacted by state and district policies that regulate how districts spend public funds to address student learning needs—including what products they can purchase. District leaders also expressed concern regarding how their systems will make decisions as they approach the ESSER funding cliff when federal grants expire, and they will return to a context of more limited resources.

As one leader shared, "...the reality of [an intervention] is that if we can't afford it, no matter how great it is, I'm not going to get it." This reality is currently muted by the availability of ESSER funding, but leaders expressed concern that when these dollars run out, they anticipate cost considerations again weighing heavily on their abil-

ity to choose one product or service over another, eliminating their ability to make decisions primarily based on the evidence of the intervention's effectiveness.

Even in a time of more significant funding due to ESSER, district leaders shared that one of the criteria they are required to use to evaluate products is cost. While cost is only one of multiple factors, some district leaders expressed that this significantly impacted their decision-making. In each case, leaders shared that the complexity associated in thinking through both the costs and benefits of a product or service was an additional challenge to decision making.

Recommendations: Budget Constraints

In the current context (spring 2023), the impending budget cliff that district leaders will face when ESSER funding expires is top of mind. This sudden fall-off of funding will impact the types of decisions district leaders will need to make in the coming years. Any evaluation network will need to take into consideration the funding picture for participating districts over the next few years. After the 2023-2024 school year, districts in most states are unlikely to have substantial funds for new programs and, in fact, may need to cut back significantly on existing investments. In short, many districts will be forced to consider what to stop doing.

We also note that in some states, districts have less local autonomy, smaller local budgets, and rely upon state-sponsored vendor contracts to drive down the price of purchased academic products and services. Here, where state education leaders are making the purchasing decisions on behalf of their districts, state leaders should be supported in using data and trials to determine the effectiveness for academic products and services.

Leadership Matters

One of the primary reasons district leaders named for the adoption or abandonment of a certain academic product was changes in leadership, primarily changes at the superintendent level. In these cases, the preferences of those with power tended to outweigh any evidence of effectiveness—or, minimally, evidence was not sufficiently visible to members of the leadership team to change direction.

District leaders cited multiple occasions when their districts had invested heavily in an academic service or product without any visible evidence of efficacy to drive the decision. As one leader shared, “A lot of times the primary decision-makers are making the decision based on either anecdotal [evidence] or their own experience. And if they're high enough up in the organization, that doesn't necessarily get questioned.” In these cases, district leaders expressed frustration with how such decisions were made, as they believed that the leader did not spend enough time understanding the feasibility of implementation in the system where they were now implementing the product or service.

Recommendations: Leadership Matters

For an evaluation network model to be effective, it is important that key decision-makers in the district play an active role in the process. This will ensure that leaders who have the most political power in their districts are brought into the use of evidence to drive decision-making processes in their systems. While the primary leader cited as making such decisions was the superintendent, multiple district leaders also named cabinet members leading the principal manager function as critical decision-makers. This investment into leadership will also enable systems to sustain shifts based on learnings from the evaluation network. District leaders shared how the lack of such executive sponsorship has, in previous networks, led to learnings being lost over time.

Political Context

The district leaders we interviewed frequently named politics within their system as a primary driver when making decisions about academic products and services to address student learning needs. This is driven by the large number of different stakeholder groups that leaders need to engage in these decisions, further adding to the

complexity of decision-making.

Describing their typical procurement processes, district leaders emphasized the importance of the various stakeholder committees they convened in coming to a decision. One leader went so far as to describe the decision-making process for a curriculum adoption as a “democratic” process, where the teachers that volunteered to test out the curriculum options came to collective decision-making based on how well they liked using the product.

District leaders must consider board direction, community input, principal buy-in, and, in many cases, teachers’ unions in these decisions. Leaders expressed the value of working across stakeholders when making decisions about academic products and services. This relates to the implementation fidelity themes above. If district leaders do not work to ensure buy-in across their districts, they will almost certainly face implementation challenges and not be able to determine the efficacy of the product or service. District leaders expressed how such change management needs significantly colored the context within which they make decisions, and ultimately why it is so complicated to use evidence to determine efficacy, even when evidence is available.

Recommendations: Political Context

As district leaders participate in evaluation networks to determine and understand evidence of effectiveness, there is an opportunity to support shared dialogue around how to navigate the complex political environment in which they make such decisions. One of the primary reasons district leaders cited as motivation to participate in an evaluation network was to have the opportunity to collaborate with other leaders experiencing similar challenges. Coupling an open dialogue with the intent of understanding and addressing political complexities with rigorous research support in an evaluation network model will provide district leaders with the opportunity to proactively use data to drive decision-making and change management needs in their individual contexts.

Balancing Intervention and Core Instruction

While this study initially focused on supplemental academic interventions, district leaders expressed the importance of balancing a focus on academic interventions with a focus on core instruction. While there was no clear preference across district leaders regarding whether a research and evaluation network should focus on core instructional products (i.e., curriculum/textbook adoptions) or academic interventions (i.e., supplemental curriculum or learning technologies), there was largely agreement that both were important, and that the connections between core products and interventions were very important in understanding the impact on student learning. The benefits of a focus on core instructional materials were primarily due to the importance of these resources for districts, as these products are the primary tool intended for Tier 1 instruction. These products tend to have common and predictable adoption cycles and typically follow a standard procurement process, making a standard evaluation cycle possible. District leaders that favored an evaluation network focused on core instructional materials cited the role of Tier 1 instruction, the significant cost of adoption, and the length of time between adoption cycles— typically 5-7 years. As one leader shared, it is important to make the right decisions regarding core curriculum because once adopted, “you’re stuck with them” for a long time. This is another case where state-level engagement is important, as some district leaders expressed their limited role in decision-making in these cases due to some states’ requirement to choose from a state-approved list of products.

Some district leaders, however, thought that there was more opportunity for learning around intervention products and services, citing how crowded the current marketplace is for academic interventions due to ESSER funding. They also named this as an area where they desire support determining what to abandon, as many systems have added new academic interventions with their ESSER funds without stopping anything. This has created overburdened systems with multiple, overlapping academic interventions occurring simultaneously, that need to now start making decisions about what to stop investing in—both due to budgetary constraints and time constraints (most importantly, students’ time).

Recommendations: Balancing Intervention and Core Instruction

While the leaders we talked to were split on whether an evaluation network should focus on core academic products or supplemental interventions, networks will need to understand how each interacts with the other. This will support district leaders' ability to make more informed decisions, as the importance of balancing use of core instructional materials and intervention products was clear in district leader responses.

Evaluation Networks

All but one district leader interviewed expressed a desire to participate in a district network for evaluation where researchers would help them, and other district leaders, incorporate evidence of effectiveness into the decisions they make about using their resources for academic products to address student learning needs. District leaders shared that this model would benefit them, and their districts, for several primary reasons: the ability to participate in and sustain a rigorous evaluation model; learning from other districts and leaders making similar decisions with similar challenges; and the political “cover” the model would provide them to make difficult decisions.

As described in the data use section, district leaders did see value in collecting and using evidence of efficacy in making decisions about academic products and services. However, collecting and analyzing these data is difficult to do. Whether this is due to the district not having the internal resources to conduct a rigorous evaluation or the complexity of the internal and external political pressures, leaders largely agreed that participating in a network evaluation model would eliminate some of these barriers. Leaders expressed that the accountability associated with participation, in addition to the expertise and formality of a research partner, can lead to more rigorous evaluation occurring in a network than if they tried to do evaluations on their own without an external support (even if they have internal capacity).

Another driver for district leaders to participate in an evaluation network is the opportunity to learn from peers engaged in similar work. This ability to learn from others both allows for a better understanding of the effectiveness of an academic product or service (i.e., does the product or service work across multiple contexts) while also providing leaders with the ability to learn from others about implementation strategies.

Finally, district leaders expressed the importance of a network for providing political cover when making difficult decisions regarding academic products and services. One leader described this benefit, despite the resources already available in their system, as follows: “So, we actually have a substantial internal evaluation...department that’s really good at this...[but] like when you’re in [a network] with a bunch of other people also evaluating something similar, it gives you a little bit more breathing room in terms of politics, which I think is really important.” This finding is particularly important in the current context, as many of the decisions district leaders face in the coming years will be focused on reducing resources due to the ESSER funding cliff.

Recommendations: Evaluation Networks

Understanding why an evaluation network would be valuable to district leaders provides the opportunity to design a network that highlights these values. If district leaders can participate in a rigorous evaluation network that holds their system accountable for participation, provides opportunities for shared data across districts, engages them in collaborative discussion to grapple with the common challenges they face when making such decisions, and provides the necessary political cover to make such decisions, there is significant opportunity to add value to participating districts. These findings also necessitate that the network design continue to be grounded in district leader input, as the complexity of the environments within which they are making such decisions will continue to impact not only their experiences, but also the experiences of students and educators whose lives are impacted each day by the decisions they make.

Feasibility of Testing Designs

Evaluations of the impact of an intervention, product, or service must compare those who receive the intervention/product/service (the treated) and those who do not (the control). This implies that to evaluate a treatment, not everyone can receive it at the same time. However, just having some students/teachers/schools that are not receiving the treatment is insufficient. To avoid biasing the results, the comparison group must also be similar to the treatment group in important ways.

District leaders were asked which of three common approaches to identifying a valid comparison group for measuring efficacy were feasible in their districts. The approaches described were using a cut score to create a treatment group, using a lottery (randomized controlled trial), or using a phased implementation.

The descriptions given for each approach were:

- (a) using a cut-score—choosing students who scored just above or below a threshold on a given test or choosing schools that were just above or below another threshold (such as percent of students receiving free or reduced-price lunch)
- (b) using a lottery—identifying eligible students or classrooms or schools and using a random number to decide which ones get the intervention first and then comparing outcomes to students not receiving the intervention
- (c) phased implementation, starting with one group of schools or grade levels within schools and then comparing how student outcomes change relative to other groups of students in the semester or year when the intervention started

Of the three approaches, the most feasible for district leaders was the phased implementation approach, followed closely by using a cut score. Only half of district leaders thought that the lottery, or randomized controlled trial, approach was feasible in their systems, with many leaders sharing concerns of equity in that lotteries do not prioritize students' need for receiving the intervention.

The most common reason district leaders gave for the higher feasibility of the phased implementation and cut score approaches were that these approaches were something their districts already did when piloting the use of new academic products or services. However, the use of these approaches, as noted above, was primarily focused on the feasibility of implementation of the new product or service, not on the efficacy of the product or service. Where district leaders did think a lottery approach could work, it depended on the district having a centralized structure that allowed the central office to tell a school whether they would implement a new product or service. Even in these cases, the use of the lottery at the class or student level was not seen as feasible due to the implementation challenges this would cause for the school or teacher.

For districts that did not think a lottery approach was feasible, the primary concern expressed was ensuring equity of resource allocation. As one district leader, who understood the research-based value of a random control trial explained: "...the risk with a randomized control trial in our work is one of two things. If it's successful, [you're asked] well why weren't you giving this to my children to begin with? Or if it's not successful, why were you wasting my children's time with this?" District leaders frequently shared that they must prioritize resources to students that need the most support. In a lottery approach, district leaders shared, the allocation of resources does not follow that logic.

Recommendations: Feasibility of Testing Designs

In order to best support district leaders with collecting and using evidence of effectiveness in the academic products and services they use, it's important to consider the political environment of the district. This will enable researchers to design approaches that honor the context of district leaders, ensuring they are able to act on the findings that come from the research. One advantage of a network is that when district leaders partner with research institutions, our respondents did express the ability to have a more research-driven approach to deciding which students would receive the intervention first.

What to Abandon

One of the primary needs district leaders reported was support in making decisions about what to abandon/stop using. Yet, one of the most difficult questions for district leaders to answer was how they currently make decisions about this. Several district leaders shared that a focus on what to stop would be the most useful benefit of participation in an evaluation network, particularly with the upcoming funding cliff that will occur when ESSER expires.

When asked why their district might abandon a product or service, district leaders shared two key reasons from their experience: a change in leadership (where the champion of the initiative is no longer able to ensure implementation), or a lack of funding. Although some district leaders did express that they would hypothetically abandon an academic product or service if there was evidence that it was not positively impacting student learning, district leaders expressed that they typically had not experienced this. The reasons given were primarily due to the difficulty of measuring impact due to the shifting agendas and budgetary priorities in their districts, and due to the implementation barriers, which clouded their understanding of whether a product or service was effective.

Even in the few cases that district leaders did share they had data indicating the product or service was not positively impacting students, they still had difficulty with abandoning it due to the political implications of such decisions. As one leader shared, “I think there’s a level of not wanting to fail...trying to change direction also acknowledges that you wasted money... and it also very likely means you wasted the most precious resource of all, which is the kids’ lives, or a portion of them. So that’s a very heavy...very serious decision, right?” This sentiment speaks to the sense of moral obligation expressed repeatedly by district leaders. They feel the weight of their decisions in a way that extends far beyond their professional efficacy—what they decide to do directly impacts the children they are responsible for serving. So, more frequently than not, it’s not until one leader leaves that the decision to abandon is possible. And in the meantime, students are impacted.

Recommendations: What to Abandon

This final theme directly relates to what district leaders expressed as their greatest, most immediate need: choosing what to stop doing. District leaders shared that this was a novel concept, and support on this front would be valued greatly. Under typical conditions, a model of an evaluation network based on identifying what products and services to discontinue was seen as helpful. Given the ESSER funding cliff district leaders will shortly face, however, this model was seen as necessary. Leaders expressed their concern that, given the typical pattern of decision-making regarding abandonment in their systems, they are not any more prepared to make such decisions under the conditions coloring the current context.

If an evaluation network is not designed to most immediately support district leaders with these complex decisions, leaders will have to make these decisions in isolation. It is therefore extremely useful to support leaders with a collective opportunity to understand what to stop doing. The structure that a research network could provide combined with the consequent political cover for leaders making hard but necessary decisions would be extremely valuable to district leaders.

Concluding Recommendations for District Engagement

In response to these findings, CEPR recommends engaging district leaders in the development of a research initiative which supports leaders in identifying what academic products and services within their current system to either modify or stop using. This will provide the necessary, just-in-time insights based on their current work, while providing a greater research base for other educators nationally.

Most evaluators see themselves persuading districts to try something new—a new curriculum or piece of soft-

ware or intervention. But when the prospect of learning requires taking a gamble on an unfamiliar new product, many leaders are understandably reticent. However, there's no need to link efficacy testing with new product adoption. There is instead high value for districts to learn about ways to implement their current interventions more effectively—e.g., with better training, or with enhanced dashboards, or with greater engagement from school-level leadership—which would not require them taking on something new. This is especially true at the present moment, as districts have launched many new initiatives in recent years and anticipate a need to cut back on academic products and services currently in use due to the expiration of federal COVID relief dollars.

There are at least three other potential benefits to testing enhancements to current interventions to determine how to modify or stop using an academic product or service, rather than simply testing new interventions: first, districts may be more willing to randomize access to additional supports—such as better training, or management dashboards or product add-ons—as long as the “control group” maintains access to the base level implementation of that product or curriculum. Second, specifically in the case of textbooks or software, the providers themselves may be more eager to cooperate and provide data to projects which involve improving implementation or testing add-ons, rather than a high-stakes, up or down efficacy test of the product itself. Third, the immediate outcomes for an implementation study—such as the amount of progress students are making on product software, or the number of minutes students are using it each week—are more continuously available throughout the school year. Of course, the ultimate goal must remain a focus on the impacts on student achievement, but those often become available only at the end of the school year, making it difficult to engage districts in an ongoing process of continuous improvement if this is the only source of data available.

When we asked one district leader for their perspective on how to collect data on efficacy of academic products or services when districts struggle to engage in rigorous research activities yet need these data to inform their decisions, the response was striking: “We’re counting on you to do that... We just don’t want you to do that on us.” While this response was somewhat flippant, the sentiment was very real. If every district wants the data, but doesn’t want the research done in their district, how will researchers provide this evidence? These recommendations provide a path forward—focus on research initiatives which support leaders in identifying what academic products and services within their current system to either modify or stop using. Then, the research isn’t being done “on” a district, but rather “with” the district.

Appendix A

The Execution Network Helping Districts Implement (or Stop) Their Current Initiatives

Study Findings

Beginning last summer, CEPR conducted a series of interviews to explore district leaders' willingness to participate in evaluation efforts for new academic products. Leaders expressed a hesitation to get involved with implementing new interventions, products, and programs given the burden of implementing the already exhaustive list of efforts they currently have in place. However, district leaders did express a strong desire to participate in an evaluation network that helped them learn how to implement their existing products more effectively, and/or to test the consequences of stopping them.

Most evaluators see themselves persuading districts to try something new—a new curriculum or piece of software or intervention. But when the prospect of learning requires taking a gamble on an unfamiliar new product, many leaders are understandably reticent. However, there's no need to link efficacy testing with new product adoption. Districts have much to learn about how to implement their current interventions more effectively—e.g., with better training, or with enhanced dashboards, or with greater engagement from school-level leadership—which would not require them taking on something new. This is especially true right now, as districts have launched a number of new initiatives this year and anticipate the expiration of federal COVID relief dollars.

There are at least three other potential benefits to testing enhancements to current interventions rather than simply testing new interventions: first, districts may be more willing to randomize access to additional supports—such as better training, or management dashboards or product add-ons—as long as the “control group” maintains access to the base level implementation of that product or curriculum. Second, specifically in the case of textbooks or software, the providers themselves may be more eager to cooperate and provide data to projects which involve improving implementation or testing add-ons, rather than a high-stakes, up or down efficacy test of the product itself. Third, the immediate outcomes for an implementation study—such as the amount of progress students are making on product software, or the number of minutes students are using it each week—are more continuously available throughout the school year. Of course, it would be important to also look for impacts on student achievement, but those often become available only at the end of the school year, making it difficult to engage districts in an ongoing process of continuous improvement.

Recommended Network Structure

Goals of the Network

CEPR is exploring launching an evaluation network for district leaders to:

1. More effectively implement high-priority academic products and interventions
2. Systematically analyze the consequences of stopping some initiatives

Target Audience

CEPR intends to recruit 5-8 districts using the following criterion. Each district should:

- Serve a minimum of 20,000 students (largely driven by sample size requirements and the transaction costs of working with many small districts)

- Predominately serve students of color and/or students experiencing poverty (reflecting our commitment to improving equity)
- Be investing in a common academic product/service with other members of the network. We will use publicly available spending data (i.e., GovSpend and Burbio) to identify groups of districts using common products or intervention types.
- Use and are willing to provide access to data from a nationally normed assessment administered at least 3x/year that measures growth in literacy and mathematics achievement (e.g., NWEA, iReady, STAR)
- Commit to participating in the evaluation network and adhering to the parameters developed by the design group

Implementation Plan

| | |
|-------------|---|
| Spring 2023 | <p>Design session with districts to address the findings from the district use of evidence study to build out the details of the evaluation network model. Questions to be answered include:</p> <ul style="list-style-type: none"> • Responsibilities of district participants, responsibilities of CEPR, and required resources (FTE and other) districts must set aside to execute • Agreement on types and scopes of interventions/products/programs to be tested • Needed timing of analyses for results to be useful in district decision-making • How districts will work together to learn from each other (e.g., sharing results, conducting certain tests together, etc.) |
|-------------|---|

| | |
|------------------|--|
| Summer-Fall 2023 | <p>Districts conduct diagnostics leveraging usage data, student outcome data, and satisfaction data to determine which academic products are in use, overlap and duplication of interventions, and which, if any, of these products to abandon.</p> <ul style="list-style-type: none"> • CEPR to provide templates and support to districts to conduct their own diagnostics. • Districts start by collecting and analyzing data in one major area (e.g., math, reading). • For the common academic product/service, CEPR creates usage reports. • District leaders report key challenges and learnings regarding the leadership moves and political support needed to effectively use student outcome data and associated metrics to effectively abandon academic products. <p>Districts and CEPR collaborate to create common implementation data schema and pilot data collection.</p> <ul style="list-style-type: none"> • CEPR creates the draft implementation data schema and incorporate feedback from districts. |
| Spring 2024 | <p>Districts develop strategies to improve implementation of their common academic product/service</p> <ul style="list-style-type: none"> • Implementation data schema rolled out to districts to pilot data collection • Districts determine new implementation strategies to test in Fall 2024. CEPR works with districts to ensure a strong evaluation design. |
| Fall 2024 | <p>Districts execute new implementation strategies</p> <ul style="list-style-type: none"> • Districts collect and monitor implementation data (using the data schema) and submit data to CEPR at the end of Fall <p>Districts undertake second round of diagnostics</p> <ul style="list-style-type: none"> • Districts focus on a different subject/practice area (e.g., SEL if the districts focused on math in Year 1) • Data schema modified as appropriate |
| Spring 2025 | <p>CEPR provides network districts with preliminary impact estimates on Fall implementation trials.</p> <p>Districts select second round of strategies to implement in Fall 2025.</p> |

