Research Use Under Pressure: State and District Implementation of the ESSA Evidence Requirements

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37364539">https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37364539</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
For my grandparents
Acknowledgments

I owe my first thanks to my committee. I feel incredibly fortunate to have had Heather Hill as my advisor and mentor. She provided steady guidance and a keen eye and attentive feedback that made my thinking and writing stronger and clearer with every iteration. Ebony Bridwell-Mitchell asked me the tough, incisive questions that pushed me to be more careful, precise, and rigorous in my logic and assumptions. Now whenever I write and read, I hear Heather and Ebony’s voices in my head asking their usual questions—and I know I am a better scholar for it. Bill Penuel ensured that my work would speak to the current field of research use; he helped me to see the value in my findings and to see myself as part of a much larger community of researchers and practitioners.

Additionally, Max Yurkofsky’s thoughtful feedback and intellectual curiosity helped push me forward when I felt stuck. My accountability group—Jerzy Wieczorek, Jen Meredith, Yang Fan, Kate Hollander, and Lindsey Madison—kept me writing and focused on my goals. And to my family and friends, near and far, thank you for the advice, the joy, and the confidence that keeps me going. To my parents, especially, thank you. I felt your love and belief in me at every step along this seven-year journey.

I am grateful to the National Center for Research in Policy and Practice (NCRPP) for introducing me to the world of research use and leading me to the questions that drove this dissertation. The work presented in the first paper was supported by NCRPP through Grant RC305C140008 from the Institute of Education Sciences, U.S. Department of Education.

I also thank the Colby College Education Program for giving me the time, space, and financial support to finish this work. I feel very fortunate to have found my new academic home among such encouraging and generous colleagues.

Finally, I am indebted to my participants for sharing their time and perspectives with me, and especially to the educators at “Arbor” and “Blake” Elementary Schools and “LEA C” who opened their doors and invited me into their conversations. Your insights shaped how I think about what it means to be a researcher and what my research is for.
Table of Contents

Abstract ....................................................................................................................... iv

Introduction ............................................................................................................... 1

Search and Selection, Implementation, or Evidence-Building: How SEAs are Approaching the ESSA Evidence Requirements ......................................................... 4

The Imposition of Instrumental Evidence Use: A Micro-Process Study ............... 57

Conclusion ................................................................................................................ 95

Appendices .............................................................................................................. 98

References ............................................................................................................... 100
Abstract

There is growing policy interest in requiring evidence-based decision-making. The premise is that, by mandating the use of research, policymakers can ensure that schools and districts invest their school improvement efforts and funding more wisely. In this dissertation, I study state and district implementation of the evidence requirements in the Every Student Succeeds Act (ESSA). In the first paper, I present three different ways that eight states approached the requirements while developing their ESSA plans: using lists of pre-sanctioned evidence-based interventions, training schools and districts to find and evaluate the research supporting potential interventions, and building local evidence of effectiveness. In the second paper, I follow one state into its first year of implementation and observe three districts attempting to use the state’s new tools and trainings. Together, these studies highlight the importance of policymakers’ interpretive processes when making sense of and adapting new policy mandates, and the challenges involved in tightening the linkages between practice and external policy pressures. They also suggest that the imposition of evidence use may at times impede the very practices and ways of thinking it aims to encourage.
Introduction

Education research has long been critiqued for having little utility for policy and practice (Kaestle, 1993), sparking many efforts to understand what research use means and how to promote it (Knott & Wildavsky, 1980; Weiss, 1977, 1980). Studies of educators’ decision-making shows that they use research in a variety of ways, from incorporating general research-based concepts to searching for relevant findings, to symbolic citations that build legitimacy and support (Coburn, Honig, & Stein, 2009; Coburn, Touré, & Yamashita, 2009; Farley-Ripple, 2012).

Meanwhile, there is growing policy interest in requiring evidence-based decision-making. The premise of evidence requirements is that, by mandating the use of research, policymakers can ensure that schools and districts invest their school improvement efforts and funding more wisely. The No Child Left Behind Act included evidence requirements in several programs, yet its impact on decision-making was mixed (Herlihy, Kemple, Bloom, Zhu, & Berlin, 2009; Weiss, Murphy-Graham, & Birkeland, 2005). Commentators pointed to problems with the specific requirements (Slavin, 2017; West, 2016) and noted that future iterations would need to be mindful of continued barriers to research use, such as limited access to relevant research and few opportunities to discuss research with colleagues (Penuel, 2015).

Nevertheless, the Every Student Succeeds Act (ESSA; 2015) includes evidence requirements based on the same premise, only now with more flexibility for states and greater specificity about the definition of evidence. What remains unclear is whether the premise of evidence
requirements holds; that is, given the difficulty and varied understandings of research use, and the many layers of interpretation and adaptation involved in policy implementation, is the imposition of research-based decision-making actually a promising means of advancing school improvement and research use?

To shed new light on this question, the papers in this dissertation explore the implementation of ESSA’s evidence requirements in two stages:

In the first paper, I rely on interviews conducted with decision-makers in eight SEAs from 2017–2018 as they helped to develop their state’s ESSA plans and prepared for their first year of implementation. I use the existing research use literature to categorize and describe three different approaches SEAs envision for the evidence requirements, from the use of lists of acceptable evidence-based interventions, to training and tools for LEAs to find their own interventions, to efforts to build local evidence. Each of these approaches has very different implications for local decision-making practice. In order to better understand why SEAs’ approaches diverged in these ways, I turned to Weick’s (1995) work and a growing movement of policy studies (e.g., Levinson, Sutton, & Winstead, 2009; Spillane, Reiser, & Reimer, 2002) that conceptualize decision-makers as individuals with “bounded rationality,” who rely on their prior knowledge and practice to filter, interpret, and adapt new policy mandates. I present patterns in the ways SEA decision-makers talked about research and how school improvement has been done in their states, and argue that their implementation approaches, while unexpected based on their degree of formal authority and research capacity, make sense as adaptations based on these perceptions and prior practices.
In the second paper, I follow one state from the previous study into its first year of implementation in 2018–2019. This state adopted an instrumental approach, meaning they provided trainings and resources to support schools and districts in finding, evaluating, and selecting research-based interventions to address their school improvement needs. By doing so, the SEA sought to address many of the barriers to instrumental research use that are frequently identified in the literature. Yet, based on observations and interviews in three districts, I show that school and district leaders still struggled to put the SEA’s tools and expectations into practice, resulting in many instances of research use that looked more symbolic than instrumental. Informed by past conceptualizations of the research-practice gap and sociocultural theory, I identify three micro-processes by which practitioners used the SEA’s tools. I argue that these micro-processes of imposed research use highlight new dimensions of the research-practice gap—namely, practitioners’ understandings of the purposes of evidence, the degree of confidence evidence should provide, and the commensurability of different forms of evidence.

Together, these papers provide a look at ESSA’s evidence requirements as they are filtered, interpreted, and put into practice. They also raise important questions about the use of policy mandates as a way to bring about greater evidence-based decision-making.
Search and Selection, Implementation, or Evidence-Building: 
How SEAs are Approaching the ESSA Evidence Requirements 

Over the last two decades, education policy has operated under the presumption that schools’ improvement efforts will be strengthened by investing in practices and programs with evidence of effectiveness. Yet what counts as sufficient and meaningful evidence of effectiveness has been under continual negotiation. Federal evidence requirements have prioritized rigorous experimental research while studies show that practitioners often rely on other forms of evidence that are more immediately relevant such as student data, anecdotes, and research syntheses (Birkeland, Murphy-Graham, & Weiss, 2005; Coburn & Talbert, 2006; Davidson, Farrell, & Penuel, 2019; Tseng, 2012). The Every Student Succeeds Act (ESSA; 2015) is the latest iteration of federal evidence requirements. By establishing multiple tiers of acceptable evidence, ESSA still privileges experimental evidence, yet also for the first time invites state and local education agencies (SEAs and LEAs) to use local evaluations and implementation evidence to fill gaps where such research does not yet exist. Therefore, as commentators have noted, this could be an opportunity for SEAs to reconceive the role of research as part of a more nuanced and inclusive sense of what evidence can be (Results for America, 2017; Kane, 2017; West, 2016).

Of course, it is by no means guaranteed that SEAs will implement ESSA’s evidence requirements in this way. New policies are interpreted and re-shaped through policymakers’ lenses of prior beliefs and practices (Coburn, 2004; Hamann & Lane, 2004; Spillane et al., 2002). Prior implementation studies suggest that SEAs may perceive different opportunities in ESSA and different approaches as more appropriate or commonsense than others,
depending, for example, on their pre-existing understandings of research and its utility for decision-making.

Drawing on documentary and interview data from eight purposively selected SEAs, this article addresses the following two research questions:

1. How are SEAs responding to ESSA’s evidence requirements?
2. How do SEA decision-makers’ beliefs and understandings about research use and school improvement relate to their divergent implementation approaches?

I begin by highlighting the key features of ESSA’s evidence requirements and what is at stake in SEAs’ implementation. Then I review the existing literature on different ways educational decision-makers use and interact with research in their decision-making, which provide a framework for organizing SEAs’ intended approaches to the evidence requirements. To answer the second research question, I draw on insights from prior implementation studies to develop a conceptual framework of SEA administrators as situated, boundedly-rational decision-makers who interpret new regulatory pressures through their relevant worldviews. Then I present findings on the three approaches SEAs in the sample have adopted and surface patterns in administrators’ perspectives that may help explain each approach. In particular, I find that the starkest differences between SEA approaches were related to administrators’ perceptions of the research base—as relevant or irrelevant to their context, and as outside of or within their control—and of the SEA’s proper role in the school improvement process. Finally, I present implications for our understanding of the role of research in supporting school improvement efforts, as well as avenues for future research.
Research Evidence and the ESSA

Evidence requirements rely on the underlying theory that schools’ improvement efforts will be more successful if decision-makers are able to identify effective interventions in which to invest their funding and efforts, and that better use of education research can help them to do so. The No Child Left Behind Act (NCLB) attached evidence requirements to several funding programs, pressing decision-makers to look for interventions with positive experimental effects. These were few and far between, which undermined the real utility and impact of the requirement (Slavin, 2017; West, 2016). Researchers observed that practitioners sometimes consulted research to inform their decisions, as envisioned by NCLB, but that more often research had limited or indirect influences on decision-making (Coburn, Honig, et al., 2009; Coburn, Touré, et al., 2009; Farley-Ripple, 2012; Weiss et al., 2005). The ESSA builds on NCLB’s evidence requirements with a few key changes; the hope is that, to a degree that NCLB failed to, ESSA will prompt educators’ understanding of, and effort to, find research-based solutions for their problems (Penuel, 2015; Slavin, 2017).

ESSA’s potential to fulfill the promise of evidence requirements is attributed to three primary changes from NCLB. First, advocates point out that ESSA’s new definition of evidence better reflects the reality of available education research, making ESSA’s evidence requirement more comprehensible and feasible. ESSA provides a four-tiered definition of evidence, ranging from “strong evidence” based on experimental studies, to “moderate evidence” from quasi-experimental studies, to “promising evidence” from correlational studies. The final tier, only usable to justify intervention selections in some funding streams, opens the door for interventions that do not yet have evidence meeting the top three tiers, as
long as decision-makers have a research-informed rationale and a plan for evaluating its implementation.

Second, advocates have also highlighted ESSA’s fourth tier of evidence as a novel opportunity for SEAs to advance new processes of evaluation and evidence-building. For example, the policy advocacy group Results for America urges SEAs and LEAs to use ESSA as the impetus to “step in and play a more active role in generating evidence” for our shared research base (Results for America & Chiefs for Change, 2018, p. 10). Kane (2017) suggests that SEAs should establish LEA networks for local pilots, or as he calls it, “a system for testing ideas” (p. 58). West (2016) agrees that if ESSA’s fourth tier is implemented well, it has the potential to add to our general knowledge base. In other words, where NCLB implementation appeared to be impeded by a lack of usable research, ESSA’s fourth tier could open the option for practitioners to try new interventions and fill gaps in the research base simultaneously (Penuel & Farrell, 2016).

Finally, and importantly for this study, ESSA devolves more flexibility and responsibility to SEAs, LEAs, and schools. In particular, SEAs have new latitude to define and determine their own approaches to the evidence requirements in their states, which could result in a wide variety of implementations. At their best, the evidence requirements could push practitioners to make better decisions about school improvement. But the autonomy given to SEAs to interpret and enact the evidence requirements also opens the possibility that they will have little effect on LEAs’ decisions, for example if research is used only to legitimize pre-existing choices, or even shift decision-making for the worse if, for example, the
pressure to meet the requirements leads practitioners to try interventions that are a poor fit for their context (Weiss, Murphy-Graham, Petrosino, & Gandhi, 2008).

Thus, as we watch ESSA’s implementation unfold across the states, we should be attentive to how SEAs choose to define compliance with the evidence requirements, how they structure LEAs’ research use process, and how they perceive the purpose and potential of the fourth tier. A few others have begun to study state responses to the evidence requirements, offering an early barometer for states’ limited uptake of evidence-related practices. For example, Results for America (2018) identified 13 ways SEAs could leverage ESSA to improve evidence use, but found that only 11 SEAs submitted plans including six or more of RFA’s promoted practices and five SEAs included none. This type of checklist analysis provides rough comparisons between SEAs that are doing more or less to promote evidence use—but it could be that SEAs understand evidence use differently. In another recent review, Dynarski (2017) looked at a sample of ten state plans, finding that, despite frequent references to “evidence-based,” the plans generally lacked specifics on what SEAs think that means, potentially masking differences in SEAs’ visions for implementation.

This study aims to fill these gaps by using an inductive approach to identify patterns in how, and perhaps why, SEAs vary in their understandings of and responses to ESSA’s evidence requirements. Its findings have implications not just for ESSA’s implementation, but for the underlying theory of requiring the use of research.
Approaches to Research Use

Many studies have endeavored to conceptualize research utilization (Farley-Ripple, May, Karpyn, Tilley, & McDonough, 2018; Weiss, 1977) and describe how practitioners typically use research (Coburn, Honig, et al., 2009; Farley-Ripple, 2012; Goertz, Barnes, Massell, Fink, & Francis, 2013). This literature has coalesced around the following framework of four common understandings of research use, ranging from substantive engagement with research findings to ceremonial and superficial uses. I use this framework to understand and organize SEAs’ responses to the ESSA evidence requirements—that is, how they envision LEAs and schools using research in their decision-making.

First, instrumental use is often policymakers’ and researchers’ presumptive or intended model of research use. It entails citing the specific research sources that support a decision—something that SEAs may want to see to confirm compliance with the evidence requirements. In instrumental use, decision-makers identify a problem and a gap in their knowledge, then seek and interpret research to generate possible solutions and weigh the pros and cons of each (Coburn, 2010; Coburn, Honig, et al., 2009; Weiss, 1977). In contrast to prior findings on the relative infrequency of instrumental use (e.g., Coburn, 2010; Coburn, Honig, et al., 2009; Farley-Ripple, 2012), a recent survey of school and LEA leaders found that the majority say they use research frequently or all of the time to inform decisions about purchasing programs, adopting curricula, and designing professional development (Penuel, Briggs, et al., 2016). On the other hand, case studies have noted practitioners’ disregard for the utility of research, preference for other forms of evidence and other concerns (e.g., stakeholder buy-in, budget constraints), and the challenging time and resource demands for instrumental use (e.g., Birkeland et al., 2005; Nelson, Leffler, & Hansen, 2009; Slavin, 2019).
In *sanctioning* use, \(^1\) the research base is still reflected in decisions, but not because decision-makers have engaged directly with the research or changed their thinking about the problem space. Instead, the process of seeking and interpreting research is conducted by a separate body, and decision-makers are required to select from the resulting list of sanctioned, research-based options (Coburn, Honig, et al., 2009; Weiss et al., 2005). This makes compliance with evidence requirements easy to confirm. Additionally, multiple NCLB-era initiatives highlight the potential of sanctioning approaches for effecting research-based decisions: studies of the Safe and Drug Free Schools (SDFS) and Reading First programs show that LEAs ended up more often adopting interventions that were on the provided lists and dropping interventions that were not (Herlihy et al., 2009; Weiss et al., 2005). One limitation of this approach is that it does not seem to change practitioners’ habits of mind for research use, nor their perceptions of what is or is not effective. For example, Weiss et al. (2005) find that some LEAs dropped D.A.R.E. in response to SDFS requirements even though they still believed it fulfilled their own aims, which were not reflected in the program’s poor evaluations. On the other side of the coin, Herlihy et al. (2009) find that schools tended to implement research-based curricula and practices regardless of Reading First’s incentives and requirements, simply because the research had already diffused and informed the field widely. Another challenge for sanctioning use is determining a credible body and process to develop the lists (Weiss et al., 2008).

---

\(^1\) Weiss first named this phenomenon “imposed use” (Birkeland et al., 2005; Weiss et al., 2005; Weiss et al., 2008) to describe the growing policy trend of evidence requirements. However, given that the “imposition” or requirement of research use could be fulfilled in multiple ways, I adopt Coburn and colleagues’ (2009) term “sanctioning” to describe the role that research plays in the decision-making process.
The first two types of research use are seemingly the best-aligned with evidence requirements, and yet **conceptual use** and **symbolic use** are often found to be much more common among educational decision-makers (Coburn, Honig, et al., 2009; Farley-Ripple, 2012). In conceptual use, research influences decision-makers’ understanding of an issue or field, but cannot necessarily be traced linearly to a resulting decision (Coburn, 2010; Coburn, Honig, et al., 2009; Coburn, Touré, et al., 2009; Farley-Ripple, 2012). Through conceptual use, research can meaningfully shift practice (Farrell & Coburn, 2016; Weiss, 1977), but because it tends to be a less conscious, more indirect process, it is difficult to cite or document. In contrast, in symbolic use, decision-makers cite research strategically in order to gain support and legitimacy, or reference research generically or selectively to support a decision after it has already been made (Coburn, Touré, et al., 2009; Farley-Ripple, 2012; Penuel, Briggs, et al., 2016). With limited time and resources, LEA and school leaders are likely to engage in symbolic use to satisfy evidence requirements (Coburn, 2010; Coburn, Touré, et al., 2009), but this does not indicate substantive engagement with research nor does it promise to change their existing practices in light of research. Conceptual use is not well-suited to the structure of evidence requirements, and symbolic use is not well-aligned with their intent—yet due to their prevalence, both types of use may still influence how SEAs interpret and respond to ESSA’s evidence requirements.

The above framework of four types of practitioners’ research use is well-established in the literature, but ESSA’s fourth tier of evidence also makes relevant a separate literature on practitioners’ participation in the **production** of research and evidence. From action research to lesson study to data-driven decision-making, advocates in this field emphasize the value for practitioners of systematically gathering data on their own practice; reflecting and making
sense of evidence, often collaboratively; and formalizing or sharing knowledge of practice (Boudett, City, & Murnane, 2013; Lewis, Perry, & Hurd, 2004; Lytle & Cochran-Smith, 1992). These models of practitioner-produced evidence are thus informed by different purposes, theories of change, and conceptions of the relationship between research and practice than the models of research use. In particular, the practitioner research field emphasizes the importance of doing research as a way of understanding one’s own work (Lytle & Cochran-Smith, 1992), whereas the research use field has traditionally presumed that research is conducted elsewhere by someone else. One exception is research-practice partnerships, which aim to link together both fields through ongoing collaborations (Penuel & Farrell, 2016; Tseng, Easton, & Supplee, 2017). Both of these areas of work may reasonably be reflected in SEAs’ responses to ESSA’s evidence requirements, depending on how SEAs interpret and integrate the fourth tier’s permission to produce evidence with the expectations of the first three tiers.

**Conceptual Framework**

In order to understand how ESSA’s evidence requirements will impact local decision-making, it is important to recognize the role SEA implementers play as interpreters and mediators of federal policy. Traditional approaches to policy studies presume that implementers make rational, strategic choices; therefore, when implementation deviates from a policy’s intent, a traditional analysis would look for problems in the incentives, resources, or clarity of the policy (Levinson et al., 2009; Spillane et al., 2002). However, policy studies have increasingly noted that this traditional approach fails to account for the way implementers understand policy mandates (Coburn, 2004; Hamann & Lane, 2004; Levinson et al., 2009; Spillane et al., 2002).
First, a traditional approach relies on an unrealistic understanding of human cognition. In reality, implementers are unable to notice and process all of the information that would result in a perfectly rational decision; rather, they have only “bounded rationality” and rely on their prior experiences and understandings to help select what to pay attention to and how to make sense of new information (Weick, 1995). For example, prior research has found that practitioners sometimes use old definitions to (mis)understand the language used in new reforms (Hill, 2001), and tend to layer reform ideas onto old ones rather than replacing them—even when that sometimes dilutes or contradicts the reform’s intention (D. K. Cohen & Hill, 2008; Spillane et al., 2002).

Additionally, a traditional approach neglects the social and situational nature of cognition. Over time, through joint work and continual negotiation, implementers tend to develop shared understandings with their colleagues. These context-specific understandings then shape what meaning people make out of new demands (Levinson et al., 2009). For example, implementers often make sense of new policy through deliberation with their peers and/or rely on trusted external experts and resources to inform their understanding (Hill, 2003; Spillane et al., 2002).

Implementers make policy implementation decisions from within this bounded, situated cognitive frame. In their study of two SEAs’ implementation of NCLB’s comprehensive school reform program, Hamann and Lane (2004) find that administrators’ differing interpretations of the program were influenced by their norms about the role of the SEA and already-established problem diagnoses and strategies. Shared understandings about
“what worked and what mattered” (p. 448) shaped how each SEA determined the most appropriate, strategic response.

Following this scholarly tradition, I conceptualize SEA administrators as situated, boundedly-rational decision-makers who rely on pre-existing, shared understandings to selectively attend to and make sense of ESSA’s evidence requirements, and to determine the most feasible and appropriate response. Below, I synthesize additional literatures on the areas of work that were most salient to SEA administrators as they interpreted the evidence requirements.

Understandings of Research and Evidence

SEA administrators made sense of ESSA’s evidence requirements by drawing on their prior understandings of what “evidence” is, whether and why research design and ongoing evaluation (as presented in ESSA’s tiers) matter, and what it looks like to make decisions “based” on evidence. Educational decision-makers use many forms of evidence beyond the research-based evidence described in ESSA (Goertz et al., 2013; Honig & Coburn, 2008). Meanwhile, the term “evidence” is increasingly used in reform discourse, sometimes meaning “data,” sometimes “research,” and sometimes both. Indeed, prior research indicates that education leaders are fairly omnivorous in their consideration of different forms of evidence—from stakeholder feedback to personal experience, student data, and research—and often use different sources and forms interchangeably (Nelson et al., 2009). Although ESSA provides an explicit definition of “evidence,” administrators’ pre-existing definitions may shape their attention to and understanding of it.
An agency’s organizational routines around research and culture of research use shape administrators’ understanding of when and how research is useful, and therefore what it looks like to arrive at a research-based decision. Recent research suggests that leaders throughout the school system value research and believe it is useful to their work (Goertz et al., 2013; Penuel, Briggs, et al., 2016). But additional research suggests that administrators do not find all forms of research equally useful, and they vary in how they weigh the value of research compared to other forms of evidence (Nelson et al., 2009; Penuel, Farrell, Allen, Toyama, & Coburn, 2016). One common concern for practitioners is finding research that they perceive to be relevant or generalizable to their local context (Birkeland et al., 2005; Nelson et al., 2009). Administrators also vary in terms of the kinds of decisions for which they routinely consult research, more regularly referring to research when discussing instruction or selecting curricula than when discussing parents or community engagement (Penuel, Briggs, et al., 2016). Research on professional learning initiatives like the National Writing Project and lesson study illustrate how continuous improvement routines can build demand for research evidence and change practitioners’ understanding of research’s utility for their work (Perry & Lewis, 2010; Stokes, 2010). Research-related perceptions and habits are likely to inform how SEA administrators envision the research-based processes and tools that would benefit LEAs.

Finally, SEAs vary in terms of the expertise and resources that are available for them to learn about and understand ESSA’s evidence requirements. SEAs possess a range of research resources, from the centrality and size of their internal research offices to their networks of internal and external sources of research knowledge (Goertz et al., 2013). Not only do these offices and networks likely shape SEA routines for what research is available and what for,
but they also may provide different recommendations and technical knowledge for responding to ESSA (Hill, 2003). We also know that school and LEA leaders have mixed levels of knowledge for how to interpret research findings (Penuel, Briggs, et al., 2016) which may affect their perceptions of research’s utility. The degree to which ESSA’s hierarchy of research rigor makes sense and is salient to SEA administrators may therefore vary based on their particular understanding of research design, and their overall perception of the requirements may be informed by their knowledge of what makes research generalizable and relevant for LEAs’ work.

**Understandings of School Improvement and LEAs**

SEA administrators’ responses to the evidence requirements were also shaped by their understanding of what SEAs can and should do to support school improvement. One dimension of this is how administrators define and value local control relative to the SEA’s involvement in decisions. For example, Zeehandelaar et al. (2015) reviewed each state’s education code and identified a wide range, between states consolidating formal authorities at the SEA and those devolving greater authority to districts. They found that 39 of the fifty states have at least partial authority to takeover schools or districts, and 18 require districts to select textbooks that have been approved by the SEA. Additionally, a multi-state study found that states are steeped in different political cultures, from the relative value they put on efficiency versus equality to their degree of decentralism, which in turn influence the types of policy instruments (e.g., mandates versus capacity-building efforts) they most commonly use to influence school improvement (Louis, Thomas, Gordon, & Febey, 2008). Differences in regulation and political culture steer SEAs toward particular understandings of how and how much they can intervene in and influence local decision-making.
Over time, SEA administrators also develop shared perceptions of their schools and districts, based on their judgments of past interactions and observations. These perceptions include how reliably others’ actions match what they say, how dependably they follow through, the degree to which they are open to and respectful of others’ input, and their competence (or lack thereof) to fulfill their responsibilities (Adams & Miskell, 2016; Bryk & Schneider, 2002). Scholars have argued that positive perceptions in these areas support productive role relationships between different groups in the school system, such as teachers, principals, and administrators, but that negative perceptions can undermine progress (Adams & Miskell, 2016; Bryk & Schneider, 2002). For example, Bryk and Schneider (2002) describe a school in Chicago where the principal failed to follow through on input from parents and teachers, and staff questioned each other’s professional competence, impeding their efforts at instructional improvement. In other words, over the course of repeated interactions and observed behavior, people build perceptions of others which then affect what future actions they believe are appropriate and likely to succeed.

**Methods**

I conducted exploratory case studies at eight SEAs using a combination of documentary and interview data and an inductive analytic approach.

**Setting and Participants**

To understand how SEA decision-makers interpreted the evidence requirements, I sought to find administrators who were involved in shaping the SEA’s ESSA plans and/or in supporting schools and LEAs in federal school improvement. I first used SEA directories
and organizational charts to identify Title I directors, lead ESSA contact persons, and directors of accountability, school improvement, federal programs, or research. In my recruitment emails, I asked participants to confirm that their work was relevant to ESSA and evidence-based interventions, or to refer me to a colleague.

For the first participant in each SEA, I aimed to include a director- or executive director-level administrator involved in school improvement or federal programs, whose immediate work therefore will be impacted by the evidence requirements. These included individuals who manage and coordinate their SEA’s system of school support and lead the implementation of federal school improvement. I also aimed to include an upper-level administrator to provide broader context about the SEA and insight into how the agency’s ESSA plans fit into its larger strategic vision. Administrator titles and divisions of responsibility varied across SEAs, but Table 1 provides a basic means of comparing the participants across SEAs. In six out of the eight sampled states, I was able to interview the SEA’s designated ESSA contact, who was involved in coordinating the SEA’s overall response to the regulations and U.S. Department of Education feedback.

Data Collection

I conducted a first round of interviews with 45 individuals at 26 SEAs. These interviews focused on common attitudes and behaviors regarding research at the agency, the state’s existing processes and priorities in school improvement, participants’ understanding and interpretations of research evidence, and their perceptions of districts involved in school improvement and turnaround. These interviews included both survey-style and open-ended questions on most topics, adapting several items from prior studies of research use (Penuel, Briggs, et al., 2016) and trust (Adams & Miskell, 2016). Interviews also included both
Table 1
Participants in Final Sample by SEA, Division, and Title at Time of Interviews (N=19)

<table>
<thead>
<tr>
<th>Participant Role</th>
<th>SEA A</th>
<th>SEA B</th>
<th>SEA C</th>
<th>SEA D</th>
<th>SEA E</th>
<th>SEA F</th>
<th>SEA G</th>
<th>SEA H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deputy, Associate, Assistant</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Superintendent/Commissioner or 1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Executive Director</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Director, Manager</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary Area of Work</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>School Support, Federal Programs,</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Accountability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESSA Contact</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Total Participants</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Note. In three states, one participant from Round 1 had retired or was unavailable for Round 2, and was replaced with a new participant, raising the total sample from 16 to 19.

questions about participants’ own work and beliefs and their perceptions of their colleagues at the SEA. For example, participants were asked to endorse various statements such as “At our SEA, there’s a widely held belief that research can help make better decisions about school improvement,” as well as to generally “describe the role of research in the way your SEA approaches its work,” and to recount the types of interactions with research they have personally had in the past year. These different question structures provided multiple angles for understanding each topic.

I used the data from these initial interviews to select a purposive sample of eight out of the original 26 SEAs. Informed by my conceptual framework of factors that may affect
administrators’ responses to the evidence requirements, I selected SEAs that varied in terms of their perception of their districts, history with local control, resources and routines for research use.

I then conducted a second round of semi-structured interviews (N=15) with the eight SEAs in the sample. These interviews focused on the state’s vision for the role of evidence-based interventions in school improvement, the state’s ESSA plan, and the decision-making process behind the plan. Among other things, participants were asked to explicate pertinent excerpts from the state plan and to rank a list of possible decision-making factors in order of their salience. I attempted to recruit the same participants for both rounds of interviews, but ended up with three new participants in round two due to staff turnover and availability. Almost all interviews were conducted over the telephone, with a small number conducted via video conference. Most interviews lasted 45-60 minutes; all were audio recorded, transcribed, and cleaned of identifying information such as the names of individuals, districts, and states prior to analysis.

I supplemented my interviews by gathering relevant documents including, but not limited to, each state’s ESSA Consolidated State Plan (as approved by the U.S. Department of Education), webinars and Powerpoints, websites, grant application templates and guidance materials, and internal decision memos. Throughout spring and summer 2018, I checked for new publicly-available documents and also sent requests to interview participants to share any new materials. As states were at different points of their implementation processes during data collection, the type and volume of documents included in this study (with the exception of the Consolidated State Plans) varied considerably across the states.
Analysis

I first analyzed my data state by state, starting with descriptive and pattern coding (Miles, Huberman, & Saldana, 2014) of SEAs’ responses to ESSA’s evidence requirements and administrators’ beliefs, understandings, and routines related to research and school improvement. Initial descriptive codes included summaries of participants’ statements (e.g., *we should continue to do what research supports*), descriptions of participants’ ways of talking (e.g., *referring to prior research findings*), and repeated phrases in participants’ own words (e.g., “*it is not something we are worried about*,” “*ESSA opened a door*”). Pattern coding involved looking across the descriptive codes and consolidating similar or related codes (see sample codes in Tables 2 and 3).

I then developed case memos about each state, considering the degree to which codes were reflected across multiple participants and data sources. I paid close attention to times when participants’ ways of talking about a topic were very similar (e.g., often participants were agreed about the nature of local control in the state), as this spoke to a more widely-held understanding. I did not discount ideas that were raised by only one participant, but I noted when and how participants’ perspectives differed; for example, if an upper-level administrator perceived research to be a much more embedded part of their work than did their colleague in federal programs, that discrepancy provided insight into the uneven presence of research in that SEA.

To answer the first research question, I used an iterative analytic approach, going back and forth between emergent code clusters and the research use literature to organize the eight SEAs into types of implementation approaches. I began by comparing data across SEAs,
focusing on the codes related to SEAs’ implementation of the evidence requirements. I used code clusters to organize the main dimensions of contrast that emerged, and I returned to the original documents and interview transcripts when necessary to refine or confirm these contrasts. For example, as shown in Table 2, one pattern that surfaced through comparative analysis was that some administrators tended to defer to ESSA’s definition of “evidence-based,” while others gave more flexible definitions of “evidence-based” practices that are conceptually supported by research. I used a matrix display to look for patterns (Miles et al., 2014), with rows representing each SEA and columns for each code cluster. I returned to the research use literature, to look for overlaps between code clusters and potential indicators of each approach (see Appendix A). Through this process, I identified three SEAs that shared codes that aligned with instrumental use, while the remaining five shared codes that aligned with sanctioning use. I divided the latter group again when I noted that four SEAs additionally fit a pattern of codes related to evidence-building.

To answer the second research question, I relied on my conceptual framework to refine and organize my initial codes into clusters related to SEA administrators’ pre-existing understandings about research and school improvement. For example, I grouped together codes reflecting SEAs’ different routines for research use, or their different perceptions of LEAs (see Table 3). I then created a new matrix display with rows representing the three SEA types identified earlier and columns representing code clusters. I looked for patterns in which codes were shared among SEAs within a type, and which distinguished them from SEAs in the other types.
### Table 2

*Sample Codes for SEA Responses to Evidence Requirements*

<table>
<thead>
<tr>
<th>Code clusters</th>
<th>Codes</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Meaning of “evidence-based” interventions</strong></td>
<td>Conceptually supported by research and evaluation findings</td>
<td>“We look for, is what they’re doing, does it seem to be sensible based on what we know from prior research?” (SEA20)</td>
</tr>
<tr>
<td></td>
<td>Defined specifically by ESSA’s tiers</td>
<td>“We reviewed all the clearinghouses […] one of the things that we said would be a challenge would be not aligning with the ESSA tiers” (SEA19)</td>
</tr>
<tr>
<td><strong>Required capacities</strong></td>
<td>Implementation skills</td>
<td>“They will have embraced a comprehensive needs assessment of the school as an ongoing process […]; they would have a good understanding of how a procedural infrastructure supports a continuous improvement process” (SEA47)</td>
</tr>
<tr>
<td></td>
<td>Using clearinghouses</td>
<td>“We tell schools, ‘here’s a link to the What Works Clearinghouse’ […] now the next training is going to be about how to utilize those repositories” (SEA14)</td>
</tr>
<tr>
<td></td>
<td>Gathering evidence for evaluation</td>
<td>“People should be looking at data […] and trying to get some sense of do the kids who got the intervention do better than other kids […] or some sort of comparative thing that helps them understand the change that they’ve made” (SEA20)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“The place where LEAs and schools are, I think, best equipped to add value is how they are specifically implementing those general practices […] And they obviously are best equipped to start to gather that evidence of effectiveness” (SEA25)</td>
</tr>
<tr>
<td>Code clusters</td>
<td>Codes</td>
<td>Examples</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Routines for research</td>
<td>Rely on external partners or translators</td>
<td>“We do use the RELs a lot; we use lots of technical assistance centers […] When we have a question we call them and say, ‘hey can you help us with this question?’” (SEA37)</td>
</tr>
<tr>
<td></td>
<td>Internal research and evaluation</td>
<td>“Our research department conducted research around the implementation and rollout of those standards […] so they conducted on-site interviews […] around the effectiveness of implementation as well as what the tie that had to improved student achievement” (SEA12)</td>
</tr>
<tr>
<td>Habits of research use</td>
<td></td>
<td>“Quite frequently we will stop and say you know, ‘let’s do some research and see what other people know about this’” (SEA06)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“We’ve built it to the point that it’s pretty ingrained, natural, like the way we do things here” (SEA20)</td>
</tr>
<tr>
<td>Beliefs about districts</td>
<td>Limited LEA competence and reliability</td>
<td>“We see in some school districts that when they do get extra money, their first inclination is to buy some new program and they don’t have the wherewithal to implement it with fidelity” (SEA34)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“I think some of their [school improvement] commitments were misguided […] and some of them don’t value school improvement” (SEA37)</td>
</tr>
<tr>
<td></td>
<td>Strong local expertise</td>
<td>“The folks that are leading the efforts at the local district level really have a better understanding than the state does about what their current needs are” (SEA12)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“I think what we’ve learned is that schools and districts know their contexts well” (SEA06)</td>
</tr>
</tbody>
</table>
Timing

Data collection for this study began in spring 2017 and concluded in fall 2018. During that time, states submitted their consolidated plans to the U.S. Department of Education for approval: 16 states and the District of Columbia submitted in spring 2017 and the remainder submitted in the fall. Many received interim feedback and were asked to revise. All fifty states, Puerto Rico, and the District of Columbia had their final plans approved by September 2018. For many states, the first cohort of schools for comprehensive support and improvement (with plans developed and reviewed under ESSA’s new evidence-based requirements) were identified in fall 2018.

Therefore, the timeframe for this study includes SEAs’ preliminary responses to ESSA’s evidence requirements that informed the development of their submitted plan and then their early efforts to clarify the actual processes, tools, and expectations for their upcoming cohorts of schools. Many participants acknowledged that their vision and plans were still tentative, with details yet to be determined and some questions still unanswered. At the time of data collection, some SEAs’ materials were still in draft or pilot form, and some had not yet been developed. Therefore, these findings should not be interpreted as representing the final shape of implementation, but rather an early look at how SEAs have differently interpreted the ESSA evidence requirements and envisioned different ways to fulfill them.

Findings

Research Question 1: How are SEAs responding to ESSA’s evidence requirements?

I found that SEAs’ implementation approaches differed in how they conceptualized LEAs’ research use. That is, the tools and processes SEAs developed and the ways administrators
explained them reflected different understandings of what forms of research LEAs should engage with, how LEAs will know what interventions qualify as “evidence-based,” and finally how LEAs should incorporate that research and those EBIs into their decision-making. SEAs’ approaches were most similar to instrumental and sanctioning use, though half of the SEAs in the sample also took up ESSA’s fourth tier and incorporated evidence-building into their approaches as well. Table 4 provides a summary of these approaches, each of which I elaborate on below, drawing on specific examples from each SEA to illustrate their defining characteristics.

Table 4
Summary of SEAs’ EBI Implementation Approaches

<table>
<thead>
<tr>
<th>SEA Type</th>
<th>Implementation approach</th>
<th>What research do LEAs engage with?</th>
<th>What does “evidence-based” mean?</th>
<th>How do LEAs decide on an EBI?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sanctioning</td>
<td>Research-based tools</td>
<td>Vetted by external organization</td>
<td>Choose from SEA’s list</td>
</tr>
<tr>
<td>2</td>
<td>Instrumental</td>
<td>Research studies and clearinghouses</td>
<td>Meeting ESSA’s evidence tiers</td>
<td>Follow guided procedure</td>
</tr>
<tr>
<td>3</td>
<td>Sanctioning &amp; Evidence-building</td>
<td>Conduct their own research/evaluations</td>
<td>Aligned with research, corroborated with local evidence</td>
<td>Use good judgment</td>
</tr>
</tbody>
</table>

Type 1: Sanctioning Use

Of all the states in the sample, SEA A’s approach most closely resembled sanctioning use of research. As described earlier, research plays a sanctioning role when decision-makers choose from pre-approved research-based options rather than interpreting research on their own. In SEA A, schools and LEAs in need of improvement will submit plans addressing a
set of effective practices that have been vetted and organized in an online platform by an external organization. Although many states subscribe to this platform, they use it quite differently. SEA A’s particular use of the online platform reflected a sanctioning approach in two important ways: first, by interpreting all practices in the platform as evidence-based rather than maintaining ESSA’s emphasis on tiers of evidence strength, and second by pre-selecting a mandatory list of practices for LEAs and schools. These steps simplified the EBI selection process, which allowed SEA A to provide more targeted assistance and will allow LEAs to focus on implementation fidelity.

SEA A’s sanctioning approach treated inclusion in the online platform as the simple definition of “evidence-based.” The platform itself provides varied “strength of evidence ratings” for each of its practices; however, SEA A’s 2018 training materials did not direct schools or LEAs to review or weigh these ratings (based on effect sizes) during their planning processes, nor did they mention ESSA’s requirements or evidence tiers based on research design. Recognizing that not all practices in the platform may meet ESSA’s new tiers, an administrator felt the solution was to speak with the platform developers to “make sure that what we have in that system […] will in fact meet that test” (SEA47)—not to train LEAs to distinguish between practices within the platform. Thus it may not be the case that SEA A administrators did not themselves perceive and understand ESSA’s new tiered definition of evidence, but that they did not see those tiers as important considerations for LEAs. Instead, for LEAs in SEA A, simply completing an improvement plan using the platform’s practices will be considered sufficient to fulfill ESSA’s evidence requirements. ESSA’s newly defined tiers of evidence have potential implications for what options schools and LEAs will be offered through the platform, but not for their decision-making process.
SEA A further turned the online platform into a sanctioning tool by identifying twelve “key” practices that will be mandatory for all low-performing schools to address. To build the list of key practices, SEA school improvement staff identified strategies they found to be most high-leverage in their professional experience and verified that they matched practices in the online platform—for example, regular communication with parents and using a tiered instructional system. In this way SEA A used the evidence requirements to impose a particular set of SEA-sanctioned, evidence-based practices. The intention here was to provide better support for LEAs. One administrator explained, “In the past we would have gone to a low-performing district and said, ‘tell us what we can do to help you.’ Now we’re flipping that. We’re walking in with, ‘Here are the evidence-based strategies that we know work. We also know [the SEA] has the support to help you get to where you need to be’” (SEA41). In SEA A’s 2018 training materials on “Selecting Your School’s [Practices],” the first step was to “Select and Assess the 12 Key [Practices]”—in other words, not really a selection at all.

For SEA A, EBI selection is and should be simple, enabling LEAs and schools to focus their attention instead on implementation, with a particular emphasis on implementation fidelity. For example, SEA A’s planning guide for schools cautioned, “A school improvement plan may be well-developed, built on relevant data, loaded with evidence-based practices to improve learning […] but the plan is only as good as the fidelity of implementation.” Administrators had plans for more periodic monitoring and coaching during the implementation process, and hoped the online platform developers would also provide more vignettes and research-based resources on “what it looks like when you’re doing it right”
(SEA47). Thus, the research that SEA A administrators wanted LEAs to engage with were models and advice on how to do EBIs rather than which EBIs to do.

**Type 2: Instrumental Use**

These next three SEAs shared a vision of helping LEAs to be “informed consumers” (SEA13) of research and thereby to make better decisions about what interventions to pursue than they have in the past. Therefore, of the states in the sample, SEAs B, C, and D most clearly exemplified instrumental use. Describing what LEAs and schools should do, administrators across these SEAs emphasized three common components: accessing a wide range of possible EBIs, knowing ESSA’s tiers of evidence, and using a thoughtful rather than “box-checking” process to evaluate and select an EBI from those options.

SEAs B, C, and D gathered existing resources and tools for LEAs and schools, aiming to provide the most expansive set of EBI options with the least onerous search process. These SEAs identified pre-existing, externally-developed repositories and clearinghouses of EBIs with a preference for breadth. For example, administrators described developing “one-stop” websites with a “robust repository of materials and resources” (SEA33), or “streamlining the approach and pointing [schools] to the clearinghouses so it’s not up to some beleaguered principal to go figure out where the evidence is” (SEA04). SEA C created a website with a growing list of over a dozen resources such as the What Works Clearinghouse, Evidence for ESSA, Results First, and Results for America. SEA D’s Title I, Part A Handbook directed users to a dozen similar sites. Rather than narrow LEAs’ attention to fewer, prioritized EBIs, these SEAs focused on providing many options. This is in clear contrast to SEA A’s efforts to narrow schools’ choices; most notably, SEA C subscribes to the same online platform as
SEA A but rather than pre-select “key” practices for schools, administrators in SEA C explained that they think the tool will be most beneficial to schools because “it has a plethora of resources” and “has all of these strategies” (SEA14).

Providing extensive options was seen as necessary for LEAs and schools to find an EBI appropriate for their needs. In SEA D’s revised school improvement handbook, they advised, “It is important to review a variety of evidence-based practices to find the best strategy to match identified school priorities.” Acknowledging the diversity of needs across their state, one SEA B administrator explained, “we are still using lists, but not quite so small or so definitive lists. In our past experience, whenever we put out a list and said, ‘these are the five approved things,’ […] it turned into true concerns raised by districts about, ‘well, but that doesn’t work for me’” (SEA04). By contrast, “pointing them to clearinghouses that already have [ESSA]-aligned interventions, […] we just thought that was a more nuanced approach and would help our districts be more able to find things that actually work for their context and their needs” (SEA04). SEA C guided schools to consider a broad definition of strategies and practices to make sure schools were “not boxing themselves in” by only considering purchasable programs (SEA14). Similarly, SEA D steered schools toward John Hattie’s (2012) work “because some of the practices that have high effect sizes like metacognition and collective efficacy, those are not age-specific, content-specific”—thus preventing the problem of “having LEAs say, ‘well what choices do we have?’” (SEA37).

Whereas SEA A treated interventions as either evidence-based (and in the platform) or not, SEAs B, C, and D explicitly instructed LEAs and schools on ESSA’s four different tiers of evidence strength. One SEA B administrator articulated the vision for LEAs as follows:
“When it’s fully implemented, they’ll know what evidence is, [that] there are levels of evidence, [and] where you can find evidence” (SEA04). Thus, implementation is not about generally being more evidence-based, but about specifically applying ESSA’s tiered definition of evidence strength to selecting interventions. In fact, the very specificity and clarity of ESSA’s requirement was what these SEAs find most promising. A SEA C administrator explained that “the way they’ve laid out levels of evidence and what’s expected with those […] I think that’s just lifted [evidence] up beyond just a catchphrase” (SEA13). These SEAs shared an understanding that one of the problems in school improvement had been schools and LEAs making decisions that either were not based in evidence (e.g., “[teachers] are resorting to looking on Pinterest for good ideas or going to Teachers Pay Teachers […] but they’re not going about it with intentional purpose” [SEA33]) or were based in anecdotal or unreliable evidence (e.g., “I mean, they didn’t quote or cite the What Works Clearinghouse […] mostly what they were quoting was whoever the publisher said the evidence came from” [SEA37]). 

SEAs C and D produced training materials that provide ESSA’s definition of evidence and SEA C’s training also walked through how to use the What Works Clearinghouse to determine an intervention’s tier of evidence.

Therefore while the search for EBIs was meant to be easy, administrators in these SEAs hoped the evaluation and selection processes would take serious consideration and time—and they produced tools and templates to structure those processes. For example, SEA B hoped to create guidance in the form of questions to help LEAs use the clearinghouses and “negotiate the world of evidence,” such as “How do you weigh out different pieces? What if there’s two studies in there that say something conflicting?” (SEA04). The aim was that LEAs will “not just [pick] a research-based strategy because, you know, ‘these are [Robert]
Marzano strategies and they’re all good, so we’re just going to pick one’; I think it’s a more thoughtful, more thorough way of trying to ascertain something that will actually work for them and move the needle” (SEA19). SEA C, which adopted the guides for identifying EBIs developed by the Florida Center for Reading Research, emphasized “walk[ing] through a process where, if you’re trying to identify an intervention for a particular problem, it lets you as a group kind of talk through several different interventions, what does the research say, what do the outcomes look like, and arrive at a consensus about what might work for your particular contexts” (SEA13). Though not all SEAs had fully implemented it by the conclusion of my data collection, they shared a vision of LEAs and schools navigating a wide range of EBIs and using a structured process to make a selection that meets ESSA’s specific tiers.

**Type 3: Sanctioning Use and Evidence-Building**

SEAs E, F, G, and H embraced the promise of evidence use more than ESSA’s specific evidence requirements. Concerned that holding tightly to ESSA’s top three tiers of evidence could force LEAs away from good, context- and culture-specific practices, their implementation approaches aimed to comply with ESSA’s requirements evidence without allowing that compliance to compromise their broader vision of LEA evidence use. In order to satisfy ESSA’s evidence requirements in the short term, these SEAs largely adopted elements of Type 1’s sanctioning approach, but, unlike Type 1, were enacting sanctioning use in ways that created space for local judgment around specific action steps. Meanwhile, their long-term ambition was to cultivate evidence-based habits of mind and evidence-building capacities in both the SEAs and LEAs in order to support continuous improvement and to generate new evidence on their locally-relevant practices—embracing ESSA’s flexible
fourth tier of evidence. This emphasis on local evidence-building as a valuable and necessary supplement to doing EBIs was what most clearly distinguishes Type 3 from the other two types.

**Short-term sanctioning use.** Much like Type 1, SEAs E, F, and H used external and internal research expertise to identify a set of practices that they were confident could meet the evidence requirements. Schools and LEAs in each state will develop improvement plans that align with the SEA’s general framework of EBIs, thereby freeing LEAs from the need to find or evaluate EBIs on their own. For example, SEA E sent their school improvement strategies to a policy research organization to analyze and “give us feedback on what is statistically sound or evidence-based” (SEA23). SEA F administrators sought to “understand[d] the evidence requirements well enough to know what of our existing research already met them […] and we were able to demonstrate convincingly that we were already doing the expectation” (SEA20). SEA H subscribes to the same externally-vetted online platform of EBIs used by SEAs A and C. In all three cases, SEA administrators felt it was important to ensure that their framework of EBIs satisfy ESSA’s high expectations.

However, administrators in SEAs F and H were quick to point out that their SEA’s sanctioned practices are intentionally broad, leaving room for LEAs to choose more specific action steps that are context-appropriate—even if those are not necessarily evidence-based.² This flexibility within “fairly general and high-level” practices (SEA25) was a key way in which SEA H administrators’ understanding of the online platform’s utility differed from

---

² SEA D’s evidence-based practices are similarly broad, but this particular intention was not explicitly articulated by SEA D administrators or documents.
those in both SEAs A and C. One SEA H administrator explained, “we are promoting our evidence-based practices [in the platform], and then the LEA […] is making decisions on how to use their funds to implement those practices in a way that perhaps has not been proven to be successful, just because there is a dearth of research on it” (SEA25). Similarly, rather than verify that LEAs’ plans strictly adhere to ESSA’s tiers, SEA F administrators planned to check, “does it seem to be sensible” (SEA20), is it a “solid component” of the school’s overall strategy (SEA20)—but “we’re not necessarily hammering home [that] everything needs to be evidence-based” (SEA46). For example, one of SEA F’s evidence-based practices is family engagement; as a SEA F administrator explained, “if a school […] wanted to use some money for family engagement strategies to have a monthly potluck […] that might not be research-based—it is going to be really hard to find studies that say potlucks lead to school turnaround—but we are not going to say to that school, ‘no, you can’t do that,’ if it is a solid component of their family engagement strategy” (SEA46).

Whereas sanctioning use for Type 1 was intended to focus LEAs’ energy on implementation with fidelity, sanctioning use for Type 3 SEAs was meant to create the space for LEAs to make context-appropriate decisions unconstrained by whether or not there is research yet to back them.

The exception within the group is SEA G, which, like Type 2, was attempting to quickly build up LEAs’ capacity for instrumental use. Regarding short term implementation, SEA G administrators shared Type 2 SEAs’ vision of LEAs and schools learning ESSA’s evidence tiers, how to “make sense of” clearinghouses (SEA48), and how to “identify appropriate […] strategies and connect the evidence from the research […] to the context of their need and
the context of their environment” (SEA38). It was in SEA G’s longer term goals that they distinguish themselves from Type 2.

**Long-term evidence building.** Type 3 SEAs were investing heavily in the theory that schools and LEAs will improve most when they are encouraged to build evidence on their implementation of local practices—both to inform their own continuous improvement processes and to establish a more relevant and rigorous knowledge base on what works. That is, administrators saw an intrinsic value in having schools and LEAs go through a process of gathering evidence on their practice, as well as an instrumental value in filling perceived gaps in the research base. Therefore, SEAs E, F, G, and H put their energy and resources into supporting action research networks, local pilots, and other evidence-building strategies using ESSA’s fourth tier of evidence.

SEA G’s plan was to gradually build an ESSA-aligned, state-specific EBI clearinghouse that includes local and SEA research. One administrator described the SEA’s vision thus: “we are putting some processes in place to evaluate those things with our research team so that when we do not have a level of evidence, we can generate one” (SEA39). In order to achieve this, SEA G planned to “build its research capacity both internally [at the SEA] and with partners to emphasize performance monitoring […] and rigorous research to meet the evidence-based requirements” (SEA G ESSA Plan). SEA G also hoped that with their regional support system, new SEA research staff, and external partners, their LEAs and schools will similarly be able to conduct research to satisfy ESSA’s expectations. SEA G planned to facilitate school- and LEA-level action research networks and rapid evaluations “all with an eye towards getting those strategies into that Level 3, if you will, Level 2, evidence strategies”
(SEA48). Their clearinghouse website told practitioners to “build up the evidence base,” because “when you evaluate the strategies being used in your districts and schools, you are contributing to the broader understanding of how evidence-based strategies work.” For SEA G, investing in a state-wide capacity-building effort was intended to enable evidence-building that would be locally-conducted and still rigorous enough to meet ESSA’s requirements.

In the long run, SEAs E, F, and H shared SEA G’s vision; however, acknowledging that rigorous research capacity was still nascent in their contexts, these SEA administrators tended to decouple the process of developing the positive habits and dispositions for evidence-building from ESSA’s specific evidence requirements. For example, SEA F produced a document that walked LEAs and schools through the process of developing a logic model and designing a program evaluation. Although the guide explained research designs that would meet ESSA’s top three tiers, administrators expressed more modest expectations for implementation: “What I would like to get to is that our districts are able and facile at using, building, and sharing evidence on their practice; […] and again, not every study they do or every analysis they do is going to meet the ESSA evidence requirements, but it is going to be better than nothing” (SEA20). Starting with evidence-building was purposeful for SEA F. As one administrator explained, “researching our own work is what gets people interested […] you start there, then you build the appetite for, ‘oh, well, you know, our research is showing that this isn’t terribly effective—is there other research that shows that a different strategy might work better?’” (SEA20). In other words, SEA F was hoping that making research relevant first would build the dispositions and curiosity necessary to seek and use rigorous research later. Similarly, SEA H planned to “require that LEAs are starting to gather the evidence that they need to establish the effectiveness of their
interventions,” but its priority was “that we are carving out a space where we can be responsive to local expertise and knowledge of need, even if that local expertise has not yet developed the capacity to document an evidence base as required most rigorously by ESSA” (SEA25). The logic was to eventually build a rigorous evidence base that will codify local expertise and knowledge, beginning first with just the habit of gathering evidence, and developing the capacity for rigor over time.

Research Question 2: How did SEA decision-makers’ beliefs and understandings about research use and school improvement relate to their divergent implementation approaches?

The second half of this paper offers an exploration of how decision-makers in different SEAs interpreted ESSA’s evidence requirements variably, through the lens of their context-specific understandings. Based on the research use and policy implementation literatures, my interviews with SEA decision-makers focused on norms, behaviors, and resources related to research; existing school improvement processes and authorities; interpretations of research evidence; and perceptions of LEAs. My analyses converged around two findings that highlighted how decision-makers relied on their local understandings and predispositions to interpret policy mandates rather than on their agencies’ formal structures and authorities.

First, I found that SEAs’ internal research capacity had a counterintuitive relationship with SEAs’ approaches. For example, we know that instrumental research use requires many supporting resources and capacities, yet while two SEAs with internal research staffs chose an instrumental approach, so did a state with very limited research capacity. What seemed to
matter to these decision-makers was their shared understanding that research is best consumed by educators but conducted externally. On the other hand, two other SEAs with very limited research capacity joined two more SEAs with well-established research offices in Type 3, sharing an understanding of evidence as primarily a way of understanding one’s own work and context.

I also found that SEAs’ formal authority to intervene in low-performing schools and districts or to influence curriculum adoptions did not match the prescriptiveness of their approaches. For example, two states with high degrees of formal authority joined two more decentralized states in Type 3, devolving both decision-making and evidence-building autonomy to the local level. What united these four states was positive perceptions of locals’ expertise and a strong value in local ownership over decision-making.

**Understandings of School Improvement and LEAs**

I use “state-centric” and “local-centric” to describe patterns in the way SEA administrators explained how local control manifests in their state and how they rationalized the appropriate degree of autonomy for LEAs in school improvement. These patterns include their formal authorities but also are shaped by their past experiences with LEAs—where tradition and public opinion has shaped their interactions, whether they have felt positive rapport with LEAs, and where LEAs have struggled or why they have found success.

**State-centric school improvement.** What set SEA A apart as “state-centric” was both its formal, regulative authorities and its political traditions. Like five other SEAs in the sample, SEA A can and has taken over low-performing districts or schools (see Table 5). Like two of
the SEAs in Type 2, it also produces an optional list of approved or recommended textbooks. Zeehandelaar et al. (2015) put SEA A on the extreme end of the centralization spectrum, based on how many authorities state education code vests at the state level. These authorities were amplified by the general tradition of centralized decision-making in the state. One administrator described how “many” of the initiatives forming SEA A’s theory of action for improvement “come from an edict from the court system, come from a piece of legislation […] , come from a State Board of Education initiative that tells us this is important, […] or could come from our state superintendent’s list of priorities” (SEA41).

Table 5
SEAs’ Formal Authority as Provided by Statute and as Enacted

<table>
<thead>
<tr>
<th>Type</th>
<th>Has statutory authority to takeover schools and/or districts</th>
<th>Has enacted takeover in past or currently</th>
<th>Produces an optional list of approved or recommended textbooks</th>
<th>Produces a mandatory list of approved textbooks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SEA A</td>
<td>SEA A</td>
<td>SEA A</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>SEA B, SEA C, SEA D</td>
<td>SEA C, SEA D</td>
<td>SEA C, SEA D</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>SEA E, SEA F, SEA G, SEA H</td>
<td>SEA E, SEA F, SEA G, SEA H</td>
<td></td>
<td>SEA E</td>
</tr>
</tbody>
</table>

Prior experience also informed SEA A administrators’ perception that LEAs need state-level direction, whether for lack of capacity or lack of follow-through. For example, one administrator shared that “it’s not like teachers don’t want to do certain things; it’s that the school districts don’t have the resources to purchase them or make good utilization of them. A part of that also does involve competence of the staff. You know, we hear and we see in
some school districts that when they do get extra money, their first inclination is to buy some new program. And they don’t have the wherewithal to implement it with fidelity” (SEA34). By contrast, when the SEA takes over a district, “we [the SEA] go in, we see where the needs are, and we actually put something in place” (SEA41). In other words, administrators perceived that the SEA has the efficacy and “wherewithal” that LEAs sometimes lack.

**Local-centric school improvement.** Although some of the other SEAs shared SEA A’s formal authorities, administrators described stronger traditions of deferring to local control. For the Type 2 SEAs, even if the SEA could or would like to do more, local control was demanded and protected by LEAs. For example, though they all had laws permitting school or district takeover, neither SEA B nor SEA D had ever exercised that authority. One administrator in SEA D pointed out, “and I don’t think that [takeover] would ever happen” (SEA37) while an administrator in SEA B noted “there was a plan to implement it this past year and the political fallout was such that […] the state] was kind of thwarted in trying to take something over” (SEA19). Thus the authority written in statute does not tell the whole story of how SEAs understand what they can and cannot, or should and should not, do. Administrators in these SEAs tended to describe their work in relationship to LEAs as more tentative and restrained. For example, SEA D produces a list of recommended textbooks, but administrators noted that “we don’t want to micromanage that” (SEA33). Similarly, one SEA C administrator explained that “[we] understand that the district at the end of the day is the one that has to implement it and we need as positive a relationship with them as we can. So for us to go in and mandate XYZ, that doesn’t sit well politically in the state, doesn’t sit well with superintendents. And so if we can draw them to the well and get them to drink, we
stand a whole lot better of a chance at getting the work done” (SEA13). It may not be their ideal, but administrators in the Type 2 SEAs were well used to structuring their work around their state’s traditions of local control.

On the other hand, all four of the Type 3 SEAs had, and had exercised, the authority to takeover schools and/or LEAs, and yet, interestingly, these administrators were more likely to talk about the value of local authority. In some cases, the SEAs had experienced political pushback as in the Type 2 SEAs, but administrators were more likely to sympathize with LEAs’ concerns. For example one senior-level administrator in SEA G recounted how LEA stakeholders “were really concerned that [school improvement under ESSA] was going to be overly prescriptive, so they wanted assurances […] that there were still lots of leeway and discretion for them to choose strategies […] based on their local context. We agree with that whole point” (SEA48). As in the Type 2 SEAs, SEA G’s LEAs actively protected their local control; yet by contrast, SEA administrators fundamentally agreed. As a colleague explained, prior state-centric approaches had ill-suited their large, diverse state, “because everyone has a different context, different need, different human capital structure, etc., so we really have to back away from that” (SEA38). When asked whether they believe LEAs need limits and parameters to guide their decision-making, both SEA H administrators were skeptical: “I guess my comment on that is that we have limited capacity here at the state, and so I am not certain that the local folks at the LEA level are not the most knowledgeable experts on making those decisions; I would probably just as a default defer to them” (SEA25). LEA pushback may constrain what SEAs G and H can dictate for their LEAs, but administrators also saw the practical value in deferring to local knowledge.
Perhaps even more surprising, Type 3 SEAs E and F had both the capacity and regulative authority to be as hands-on as SEA A, and yet administrators’ interpretations of what has worked or not worked in the past, and why, seemed to be more powerful influences on how they see their appropriate role with LEAs. For example, SEA F had “robust laws here, in particular the one that helps us in circumstances where schools are not performing well” (SEA06) as well as the human resources to “go pretty deep” with their low-performing schools (SEA46). However, outside of the few schools and LEAs under takeover, SEA F conceived of its role and authority as “enabling schools to get the good information that they need and helping them with critical friend support” because “I think what we have learned is that schools and districts know their contexts well” (SEA06). SEA E’s formal authority was even broader, bolstered by an “entire legislative environment [that] is very controlling about what districts can and cannot do” (SEA12). Its staffing capacity dwarfed that of its mostly small and rural LEAs, and administrators acknowledged that this means the SEA had been very involved in LEA operations. Yet SEA E administrators wanted to make an effort to leave more control in local hands, sharing an “understanding that the folks that are leading the efforts at the local district level really have a better understanding than the state does about what their current needs are” (SEA12). In other words, these administrators saw the SEA as providing extra capacity, but LEAs as fully competent decision-makers.

SEA administrators’ understanding of decision-making was, of course, shaped by their official responsibilities and authorities, yet was also strongly shaped by experience and perception. SEA F administrators “have learned” through past efforts in school improvement and turnaround to trust the competence of their LEAs and schools. SEA G
administrators perceived local discretion as necessary to respond to their LEAs’ varied needs. Remembering prior “fallout” and what “sits well” with LEAs, the Type 2 SEAs saw delegating authority as the most politically and practically expedient route. Meanwhile, for a host of reasons, SEA A administrators had come to the opposite understanding.

**Understandings of Research and Evidence**

I use “external” and “internal” to describe patterns in the way SEA administrators explain their routines and beliefs about research, from what kinds of questions they ask, to where they go for answers, to what they think research is and is for. Administrators across the board positively endorsed statements on the utility and credibility of research. Where they differed was whether their routines and beliefs supported research as a source of externally-originating knowledge or an internally-driven process. I argue that administrators’ understanding of the role and utility of research in their work at the SEA was related to their perception of how LEAs could best benefit from research in their work.

**External research.** Administrators in SEAs A, B, C, and D generally talked about research in ways that match the linear, uni-directionality of instrumental research use, where knowledge originates out in the research community and then is transferred into the SEA. A major contributing factor, understandably, seemed to be that (at the time of data collection) three of these SEAs did not have their own research offices or dedicated staff. This lack of internal resources for research did not mean that administrators therefore had less access or engagement with research, but simply that they looked outward instead. And it may be that habitually looking outward for answers shaped administrators’ understanding of research and their relationship to it. In particular, in the ways they described their interactions with
research, administrators in these SEAs assumed the applicability of outside research to their questions and decisions, and they focused on the value of research as a product (which can be conducted by anyone) rather than a process.

Every administrator in these four SEAs was able to point to multiple external partners or organizations whom they are able to access easily and often. And, most often, the kind of help they wanted and the questions they asked took the form of research-based training, tools, and literature reviews—all products that could quickly and directly be applied to their work. For example, when describing their research activities from the past year, one administrator in SEA C spoke about “relying heavily” on research clearinghouses, using the comprehensive center “pretty extensively,” and “reach[ing] out” to other external organizations “to pull from other states what might be working in other places or what’s going on” (SEA13). When asked what they would do with a research question, a SEA D administrator gave the example of how, when they wanted to revise their improvement system, they reached out to their comprehensive center who “actually conducted a lit review on our behalf […] and then we took that information and developed our school improvement needs assessment” (SEA33). And “then we had that research base that [the comprehensive center] had pulled together for us to support what we were proposing” (SEA33). Although later, when talking specifically about ESSA and EBIs, some administrators were skeptical that an intervention that works in one district could work as well in a different context, here, in their own routine work with research, they emphasized the quick application and transferability of research-based syntheses and best practices.
SEA B’s experience suggests that seeing research as a source of externally-originating knowledge is not simply about having or not having a research office, but also is shaped by how those offices are used and understood. Unlike the other SEAs, SEA B had a research office at times in the past, but a lack of political support for research made it difficult to sustain, with inconsistent funding and unfilled positions. Administrators remembered the office with frustration: “I have never been able to get the research questions answered that I have wanted to; like two years ago they started a project for me and didn’t have the capacity to finish it” (SEA19). In contrast, this administrator spoke glowingly about the support and attentiveness of their comprehensive center staff: “she shares things with us all the time and if we have questions, she is always willing to find the research we need” (SEA19). At the time of data collection, SEA B was in the middle of “basically offshore[ing] some of that [research] work that we’ve been trying to do in-house” (SEA04). An upper-level administrator who had watched the SEA’s research office rise and fall “for a variety reasons, leadership changes … a lot of things happened,” concluded succinctly that “the power of the research community is in the research community” (SEA04). Therefore even though they had a research office, SEA B administrators seemed to have arrived at similar understandings to SEAs C, D, and A: first, that for practical reasons, research is more reliably and effectively done not in the SEA; and second, that research work is “offshore”-able.

**Internal research.** SEAs E, F, G, and H represented disparate degrees of research capacity, but their administrators shared a certain skepticism of the utility and relevance of outside research. With the exception of SEA H, the SEAs also found ways to make research an ongoing and integrated part of their work. These research routines and attitudes seemed to go hand in hand.
In contrast to the administrators described above, administrators in the Type 3 SEAs described external research with more caveats about when and to what degree it is applicable to their decisions, often pointing to examples of when research failed their needs. For example, an administrator in SEA F explained that the external research base trails behind their needs:

“We do like to use as good research as we can [… but] sometimes we are so much on the edge of change, things that have not been done before. So we have to, with our eyes wide open, accept that we may not have perfect data to move forward. [But] for us to wait until there are these research studies that are well-designed and all of that would leave a lot of schools behind […] if we did not try to carefully implement some [other] things that we think may be useful.” (SEA06)

Other Type 3 SEAs saw the research base as rarely applicable to their contexts. For example, an administrator in SEA E recounted their experience with the research-backed School Improvement Grant models, where “a lot of those turnaround principles were really about big city schools and it is not practical in a rural state like [ours …] so we got a lot of pushback […] about ‘they don’t know the environment or the culture that we are in’” (SEA23). An SEA H administrator shared a similar story with Success For All: “if a school or a district [here] adopts that program, by the nature of their size, they have to fine-tune it and implement it a little bit different than it’s been implemented back in Baltimore, right? And so that is a concern […] because these small schools, multi-level classrooms are just a different beast that has often not been researched explicitly” (SEA25). Although these administrators also acknowledged help from their comprehensive centers or clearinghouses, compared to SEAs A, B, C, and D, their dominant narratives about external research highlight obstacles and limitations rather than ease of use.
Having an internal research office supported and was supported by these SEAs’ understanding that conducting new research and setting one’s own research agenda should be closely tied to an agency’s practice and decision-making. In SEA E, the research office was small and their processes were ad hoc, yet its administrators described regularly getting help with research-based answers to support the state’s decision-making. They recounted times their research office developed literature reviews and “just things for the Board to look at when they were trying to transform how they wanted to think about professional learning” (SEA23), and conducted a mixed methods study of how LEAs were implementing new standards (SEA12). SEA G had only recently reinvigorated its research office, but its aims included “build[ing] an internal community of practice with those people who are interested in research” (SEA39) and “build[ing] a better learning agenda” (SEA48). SEA F’s research office fulfilled many of the same functions, only from a more established position with more staff, clear routines, and an annual research agenda. Every year, staff from SEA F’s research office “go to the program staff and […] say, essentially, ‘what is it we need to learn this year to help you do your job better?’” (SEA20). This research routine had been “built to the point that it is pretty ingrained, natural, like the way we do things here” (SEA20). This administrator also noted that whereas outside the SEA, “I think people also get very caught up on impact analysis, […] really, when you talk to my program staff […] probably 80 percent of the questions that they raise are not impact analyses” (SEA20). Therefore, having an internal research office means that SEA staff were able to ask different questions for which they did not see answers externally in existing research. Although these offices had different capacities, their central presence and routine use within the SEAs contributed to an understanding of research ideally as an internally-driven process.
The outlier here was SEA H, which had no internal research capacity at present—and yet administrators still shared a perception of research as something that should and could be involved in. Administrators cited “time limitations” (SEA25) and budget cuts with a “bare bones” staff (SEA29) as obstacles for the agency’s engagement in research. In their routines, SEA H administrators sounded a lot like the Type 2 SEAs, relying on external supports like their comprehensive center “because we can take those questions to them and they have staff you know who can dig in and do that kind of background research on you know different policy options that we may not have the time nor capacity to do” (SEA25). And yet, when they talked hypothetically, these administrators still shared a perception that research could be better done within the SEA. Describing how to address the concern that research did not often apply, one administrator suggested, “well, let’s find a study or find something, try it in these smaller schools, run a study, do the experiment here—then we have data that backs up our specific circumstances because we are not the same” (SEA29). A colleague agreed that, ideally, “we would undertake more of a research function to look at selected practices in the state” (SEA25). Like SEA B above, SEA H again complicates the relationship between having (or not having) resources for research and having a particular understanding of what research should be for and by whom.

**How School Improvement and Research Intersect**

Figure 1 summarizes the common understandings about school improvement and research use that were shared by decision-makers in each SEA.

In Type 1, SEA A administrators had long relied on others’ research to guide their work and decisions. These traditions supported their positive beliefs that external research could and
had identified best practices that could and should be replicated in their LEAs. Yet while the SEA has regularly sought and used research instrumentally, administrators envisioned their LEAs engaging in sanctioning use instead. When SEA A administrators talked about school improvement, the availability or appropriateness of EBIs was not a problem to them; the problem was LEAs’ ability to follow through and implement them with fidelity. Therefore, a sanctioning approach appeared to those decision-makers to be a commonsense way to efficiently fulfill the evidence requirements and help focus the SEA’s hands-on support on replicating EBIs’ promising results.

**Figure 1**

*Summary of SEA Decision-makers’ Shared Understandings*

<table>
<thead>
<tr>
<th>School improvement direction-setting should be…</th>
<th>State-centric</th>
<th>Local-centric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Externally</td>
<td>SEA A</td>
<td>SEAs B, C, D</td>
</tr>
<tr>
<td>(Type 1 Sanctioning)</td>
<td>(Type 2 Instrumental)</td>
<td></td>
</tr>
<tr>
<td>Internally</td>
<td>SEAs E, F, G, H</td>
<td></td>
</tr>
<tr>
<td>(Type 3 Evidence-building)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Like Type 1, Type 2 SEA routines involved deferring to external research experts for answers to their questions, but what set them apart was that Type 2 administrators also found themselves deferring to strongly-asserted local control. Administrators’ talk about their state context and work with LEAs frequently reflected a need for caution and restraint when exercising SEA authority. Type 2 SEAs’ instrumental approach to EBIs makes sense in
light of these characteristics. With decisions made at the local level and the evidence to inform better decisions primarily developed externally by research and intermediary organizations, the appropriate SEA role was neither to make decisions nor to create evidence, but instead to make a process that connects local decision-making with external research.

Finally, like Type 2, Type 3 SEA administrators saw LEAs as the necessary ultimate decision-makers. Yet even while administrators in these SEAs valued research evidence overall, they were more focused than their Type 2 colleagues on the limitations of external research. They tended to describe their LEAs as doing good work but the research community as lagging and lacking relevant evidence. By contrast, administrators spoke readily of the benefits of generating and using their own research to support continuous improvement on locally-appropriate or innovative strategies. Thus a loose sanctioning approach combined with evidence-building may have surfaced as the way to help LEAs reap the most benefits of research-based decision-making.

**Discussion**

**Summary of Main Findings**

How SEAs implement ESSA’s new evidence requirements will reshape the policy context within which LEAs use research (or not) to make consequential decisions about school improvement. In this exploratory study, I found that SEAs held different conceptions of what it means to fulfill these requirements. In Type 1, SEA A administrators adopted a straightforward understanding of EBIs: a trusted source has determined what practices are evidence-based, and the SEA will require that LEAs implement a pre-selected list of
practices. Type 2 SEAs honed in instead on ESSA’s multi-tiered definition of “evidence-based” and will train LEAs to find, evaluate, and select EBIs accordingly. For these SEAs, interventions were no longer simply evidence-based or not evidence-based, but were supported by bodies of research with different levels of confidence, and it will be up to LEAs to weigh those strengths in a structured process of instrumental use. Finally, Type 3 SEAs conveyed perhaps the most nuanced sense of EBIs. Though strongly invested in the spirit of evidence-based decision-making, administrators in these SEAs were more skeptical of the letter of ESSA’s requirements based on what research is available and what LEAs are ready for. Therefore, they saw a sanctioning approach as a way to technically comply with the requirements while building capacity and original evidence.

This study also found that SEAs’ approaches were connected to SEA administrators’ different understandings and routines regarding research and decision-making authority, which shaped what appeared to be the most appropriate and feasible responses. First, I found that SEA administrators were accustomed to different patterns of state- or local-centric decision-making in school improvement. Type 1 administrators were more likely to describe the SEA as the primary decision-maker and direction-setter. This orientation informed their sanctioning approach: LEAs will not directly engage with research to select EBIs because those selections will be made for them. The other SEAs’ administrators instead perceived their LEAs as the primary leaders and decision-makers for school improvement. Second, what set these other SEAs apart was that Type 2 administrators described external research as providing the expertise LEAs need for their decisions, whereas Type 3 administrators cited meaningful lessons from their SEA’s internal research as compared to limited relevance from external research. These different understandings of
their relationship to research made an instrumental approach more sensible to Type 2 administrators and a sanctioning and evidence-building approach more appropriate to Type 3 administrators.

Implications

This study’s insight into the prior understandings that SEA decision-makers draw on when determining their implementation approaches should make us question the underlying theory of evidence requirements. Evidence requirements presume that the education research base contains findings that are relevant and useful to the problems and decisions involved in school improvement. ESSA’s evidence tiers steer practitioners toward more rigorous studies with experimental designs, reflecting the understanding that greater research rigor should give practitioners greater confidence in the findings. The administrators in SEAs A, B, C, and D seemed to agree: research quality matters, is more efficiently achieved by external, full-time researchers, and research findings can be immediately applied. Yet half of the administrators in the study indicated the opposite—that they would put the greatest confidence in locally-generated evidence, even when that involved temporarily putting aside ESSA’s expectations for rigorous design. There were many intertwined reasons for this perception: administrators vouched for the local judgment and expertise of their district and school leaders; they perceived the current research base to have limited generalizability for particular demographics and contexts; and for some administrators, internal research routines had become part of the SEA’s own learning process. What is interesting is that, when asked directly, the SEA administrators in this study uniformly agreed or strongly agreed that research is trustworthy and can help make better decisions about school
improvement. The variable understandings of research, researchers, and the relationship between relevance and rigor are worth exploring further.

Another question for the theory of how evidence is intended to support improvement arises from the way SEA A adapted the mandate to their existing understandings about school improvement and the role of the SEA. Administrators in SEA A saw improvement initiatives and best practices being determined either by state actors or by organizations creating research-based tools; districts were not seen as key decision-makers and therefore the evidence requirements were seen to have little relevance to their work. Weiss (1980) noted that officials struggled to describe instances of research use in part because they struggled to identify as decision-makers. With decisions coming from above them, made with many others’ input, and/or made incrementally, they rarely saw themselves owning a specific decision or consciously using research for it; rather, research concepts diffused into their working knowledge in less perceptible ways. Indeed, later research corroborates her sense that conceptual use is quite common, if under-acknowledged (Coburn, 2010; Farrell & Coburn, 2016). SEA administrators adapt policy mandates to fit the way work is done in their context; if the evidence requirements are a poor fit for the way school improvement decisions are actually made, then, as in SEA A, they may have little influence.

Finally, Type 3’s approach raises the question whether engaging in research use most substantively requires, at least for a while, not doing the evidence requirements. Type 3’s combination of sanctioning use and evidence building is both novel and promising in the way it positions practitioners as both consumers and producers of research. The SEAs in this group created sanctioning lists that more closely resembled the kinds of research-based
frameworks and broad priorities that characterize conceptual use (e.g., a culture of high expectations, instructional leadership), than a menu of proven-effective programs. On one hand, taking a more conceptual approach and pushing practitioners to use a research-based framework might lead to more substantive changes in thinking. But on the other hand, by considering ESSA’s requirements satisfied at this broad level, and allowing practitioners to decide their specific actions without regard for their evidence base, these SEAs are sidestepping ESSA’s intended use. Similarly, by taking up ESSA’s fourth tier, the Type 3 SEAs reflect commentators’ high hopes, attempting to generate new evidence on context-specific practices, and help practitioners learn from and be persuaded by evidence on their own and peers’ practice (Kane, 2017; West, 2016). At the same time, to engage in evidence-building, the SEAs have also temporarily put aside ESSA’s expectations for rigor. As one SEA F administrator put it, they are “less interested” in meeting ESSA’s evidence requirements than in “our aspiration is that the whole state gets better at using evidence” (SEA20). These SEAs suggest that requirement of research use may get in the way of effective, substantive research.

**Limitations and Directions for Future Research**

There are three major limitations to what this exploratory study can tell us. First, the sample of SEAs is small and not representative. The eight SEAs were selected to reflect a range of ESSA implementation approaches present in the initial sample of 26 states, but it should not be assumed that the three types are an exhaustive representation of all approaches across the country. Additionally, the relative prevalence of the three types in the sample are not necessarily reflective of trends in the broader population of states.
The small sample of decision-makers in the study also means that the patterns picked up in the analysis are fairly noisy. For example, the administrators in SEAs B through H are grouped together as sharing a local-centric understanding of decision-making, yet there was considerable intra-group variation in whether this orientation was more related to LEAs asserting their control or SEAs valuing LEA expertise and leadership. Similarly, the categorization of internal understandings of research is a consolidation of several more specific themes around relevance, capacity, routines, and perceptions of “the research.” There are also plausible alternate factors that emerged in some administrators’ comments but were not measured systematically in this study, such as the possibility that rurality or student demographics (particularly groups that are the subject of fewer studies) were the root of administrators’ perceptions of the relevance of external research. Future studies with greater numbers of participants would help to tease these important differences apart and perhaps detect new patterns not capture in this study.

Finally, this study’s implications are limited by its timeframe. In fall 2018, when data collection concluded, many SEAs were still very early in their implementation, and some of the findings are based only on administrators’ hopes and projections. Therefore, additional research could address whether and why SEAs’ approaches change over time as they transition from idea to reality. It could be that in practice, these three approaches in fact look much more similar, as SEAs find some elements impractical to implement and drop them, or look to other states and decide to mimic their approaches. Or it could be that in practice, these three approaches splinter into many more as SEAs adapt these principles to the unique needs, capacities, and feedback of their agencies and districts. SEAs are also constantly in flux; at the time of data collection, multiple SEAs were in the process of getting new
research offices approved, funded, and staffed, or were launching new collaborations with outside research centers—all of which could dramatically change their relationships with research and sense of appropriate EBI work.

In addition, as states begin to implement their approaches, future research should explore how each approach plays out at the LEA level. These findings have discussed the different aspects of ESSA that are emphasized in each approach, but we will need to look at LEAs’ experiences to see whether and how these approaches differ in learnability and feasibility, and whether and how they shape the decisions LEAs make. For example, SEA E’s approach banks on the theory that building evidence on local practice will foster LEAs’ curiosity about external research evidence—but how does that manifest in practice? The Type 2 SEAs plan to train LEAs about ESSA’s tiers of evidence and how to evaluate research accordingly—how will LEAs fit these new procedures into their existing decision-making practices?

Over the past decade researchers and practitioners alike have innovated and studied different ways of using research to inform school improvement decisions. This study suggests that ESSA’s evidence requirements, as divergently implemented across the states, pose a new and important opportunity to see whether and how local decision-making practice responds to different visions of research use.
The Imposition of Instrumental Evidence Use: A Micro-Process Study

In his essay on the “awful reputation of education research,” historian Carl Kaestle (1993) described the widespread perception that education research had little connection to practice. Researchers and policymakers pointed to problems in both the dissemination of and demand for research. Over the past two decades, federal policymakers have attempted to bridge this gap between research and practice through the imposition of evidence requirements. These requirements pressured practitioners to invest resources in programs and practices that were supported by research evidence. They have been met with mixed success. Educators continued to use a wide range of non-research evidence in a similarly wide range of manners that went beyond the kind of evidence use sanctioned by policy (Coburn, Honig, et al., 2009; Coburn, Touré, et al., 2009; Honig & Coburn, 2008; Weiss et al., 2005; Weiss et al., 2008). In some cases, districts changed program adoptions in response to requirements, but not because they actually consulted or believed the evaluation evidence (Birkeland et al., 2005; Weiss et al., 2005). In other cases, schools adopted research-based practices and curricula, but individual teachers’ uptake of those practices varied (Herlihy et al., 2009; Penuel, Frank, Sun, Kim, & Singleton, 2013).

This policy trend has been accompanied by a growing body of literature aiming to conceptualize the research-practice gap and better understand why research is so rarely used to inform decision-making. This literature has pushed past simple problem framings around supply and demand, considering instead what a supply of useable research actually looks like, and what capacities or organizational characteristics motivate and enable meaningful use. Practitioners need research relevant to their particular problems of practice, presented
accessibly, and supplied by trusted intermediaries or research brokers (Conaway, 2020; Farley-Ripple et al., 2018; Penuel, Farrell, et al., 2016). They also need time, routines, and the skills to interpret research and draw out its implications for their practice (Coburn, Honig, et al., 2009; Coburn, Touré, et al., 2009; Knott & Wildavsky, 1980; Stein & Coburn, 2010).

This last component—the skills or understandings that are necessary to turn research evidence into actionable knowledge—is widely acknowledged to be crucial for research use, but is still under-elaborated as a construct (cf. Corcoran, Fuhrman, & Belcher, 2001). Some have suggested that practitioners should be familiar with principles of study design, causation and correlation, and sources of bias in order to make and evaluate appropriate inferences (Dynarski, 2008; Penuel, Briggs, et al., 2016). Yet others have theorized that using research relies on more than just knowledge. For example, in addition to knowing common threats to validity, practitioners may need to be able to reconcile different conceptions of validity and conclusiveness from the worlds of practice and research (Farley-Ripple et al., 2018). As tools proliferate that pre-evaluate and summarize research, we need to better understand what additional understandings and processes practitioners rely on when they use research to inform decisions.

The Every Student Succeeds Act (ESSA; 2015) provided an opportunity to observe schools’ and districts’ research use in practice. ESSA includes a new iteration of evidence requirements: schools and districts need to select school improvement interventions that meet a particular set of criteria for evidence of effectiveness. This imposition allowed me to capture research use as it occurred, while schools and districts developed their school improvement plans.
Additionally, the state education agency (SEA) in this study prescribed routines and provided abundant resources to help practitioners meet ESSA’s evidence requirements. These implementation tools helped to bridge many of the better-understood components of the research-practice gap. Identifying patterns in the ways practitioners still struggled to use research as intended by policymakers, therefore, illuminates new aspects of the gap.

I conducted observations and interviews with practitioners engaged in decision-making meetings at three different districts and focused my analysis on the micro-processes (Little, 2011) of research use. What I found was that practitioners successfully selected programs and strategies that met ESSA’s evidence requirements, and that they used their SEA’s tools and routines while doing so. Yet their actual practice of using research was not so straightforward. At times practitioners delayed and delegated research use, used research without being convinced by it, and let research conclusions completely supplant other considerations. I argue that each of these micro-processes point to ways in which differing conceptions of evidence—and not being able to reconcile them—maintain a gap between research and practice.

I begin by contextualizing the SEA’s efforts in the recent history of federal evidence requirements and the challenges of imposing research use. I review the literature on the research-practice gap and argue that as tools and supports for research use grow over time, we need micro-process studies to understand what aspects of the gap persist in practice. Finally, I present my findings on three micro-processes of imposed research use in the
districts and close with a discussion of their broader implications for our understanding of connecting research and practice.

**Efforts to Impose Research Use**

Policymakers’ presumptive ideal model of research use is instrumental. To use research instrumentally, practitioners first identify their problem or question, then find relevant research and use that evidence to weigh and select from their options (Coburn, Honig, et al., 2009). The logic of imposing instrumental use is that schools’ and districts’ outcomes will improve when they select programs and practices that have demonstrated their effectiveness, and stop or avoid those that have not.

Yet, past efforts to impose instrumental evidence use have ironically tended to incentivize or yield other forms of use. The No Child Left Behind Act (NCLB) included numerous requirements for adopted programs and practices to be based on “scientifically based research.” Multiple studies from the NCLB era show that, in practice, instrumental evidence use was not particularly common (Coburn, 2010; Coburn, Honig, et al., 2009; Farley-Ripple, 2012). More frequent practices involved post hoc and generic references to what “research says,” especially when practitioners faced pressure to deliver evidence-based decisions with limited time and resources (Coburn, Touré, et al., 2009). Studies also found that practitioners often relied on forms of evidence outside of those promoted and sanctioned by the policy: for example, practitioners based decisions on their professional judgment and informal context-specific knowledge (Coburn & Talbert, 2006; Honig & Coburn, 2008). Certain NCLB initiatives, such as the Safe and Drug Free Schools and Reading First programs, went a step further and provided lists of pre-approved interventions. Some research on these
initiatives suggest that districts tended to adopt the pre-approved interventions and drop interventions that were not on the lists (Herlihy et al., 2009; Weiss et al., 2005), though other research cautions that classroom-level implementation still varied (Penuel et al., 2013). Even when the end decisions may have looked like the results of instrumental research use, researchers noted that the process resembled a bureaucratic exercise (Weiss et al., 2008).

For those frustrated by the unfulfilled promise of NCLB’s evidence requirements, ESSA represents an opportunity to get it right. Like NCLB, ESSA continues to require research use in significant decisions. The U. S. Department of Education’s (2016) non-regulatory guidance, for example, encourages schools and districts to “determine the interventions that will best serve their needs […] by using rigorous and relevant [research] evidence and assessing the local capacity to implement the intervention.” To identify which interventions are supported by “rigorous” research evidence, ESSA defines four tiers of allowable research evidence: the strongest evidence is based on experimental studies, while “moderate” and “promising” evidence are based on quasi-experimental and correlational studies, respectively. A fourth tier of evidence permits interventions that are accompanied with a strong rationale and a plan to evaluate implementation. “Interventions” include programs and curricula as well as instructional and organizational strategies for improving student outcomes.

The State’s Approach

The SEA in this study focused its implementation of ESSA’s evidence requirements on the law’s elaborated definition of “evidence-based interventions” and providing implementation tools and resources for schools and districts to engage in instrumental use. They provided links to and optional trainings on how to navigate the research clearinghouses such as
Evidence for ESSA and the What Works Clearinghouse (WWC), and how to evaluate the research evidence according to ESSA’s four tiers. The WWC has syntheses of effectiveness studies on hundreds of programs and practices (Institute of Educational Sciences, 2020), while Evidence for ESSA lists over one hundred programs that meet ESSA’s evidence standards (Center for Research and Reform in Education, 2020).

The SEA also provided schools and districts with a handbook, rubrics, and templates that elaborated on and document each step of the school improvement planning process. The improvement plan template required documentation of the research evidence that supported their choice of programs or strategies. Finally, external consultants were vetted, contracted, and trained on the evidence requirements in order to provide objective expertise and facilitate the improvement planning process with schools. In all these ways, the SEA clarifies and elaborates a particular understanding of what counts as research evidence and what instrumental use looks like.

The Research-Practice Gap

Over time, researchers have identified a set of diverse factors that explain why it is difficult to connect research and practice. Prominent early explanations focused on problems with the supply of and demand for research. On the supply side, research needs to be relevant and made available in usable form. For example, Knott and Wildavsky (1980) noted that widespread calls to increase the dissemination of research relied on the assumption that either not enough research existed or that practitioners were not aware of existing research. They argued instead that practitioners do not need more dissemination, but better dissemination; research knowledge needed to be filtered and translated for use. Later studies
continued to find that problems with the format, timeliness, accessibility, and relevance of research impeded practitioners’ use (Broekkamp & van Hout-Wolters, 2007; Coburn, Honig, et al., 2009; Honig & Coburn, 2008; Nelson et al., 2009; Penuel, Briggs, et al., 2016).

On the demand side, practitioners need to seek and attend to research. Some studies of research use have identified a lack of motivation or desire as a major barrier to its use (Nelson et al., 2009), though a more recent survey of school and district leaders suggested that the majority shared very positive beliefs about the value of research and frequently looked for research to help with problems and decisions (Penuel, Briggs, et al., 2016). Weiss et al. (2008) also argued that federal evidence requirements “effectively overcame [policymakers’] inattention to research and their unwillingness to take it seriously in making decisions” (p. 43), by making research use a condition of receiving federal funds. Supply and demand certainly matter for bridging the research-practice gap. However, a proliferation of research clearinghouses and use-oriented research tools, as well as continued regulatory pressure mean that practitioners are increasingly likely to encounter and try to use research.

Delving a step further into the research-practice gap, there is a growing consensus that routines and relationships are key to structuring and facilitating meaningful interactions with research. First, routines can create opportunities for research to enter the conversation, such as if districts follow a particular decision-making process or if teachers use a continuous improvement protocol that include steps for consulting research (Corcoran et al., 2001; Penuel, Briggs, et al., 2016; Perry & Lewis, 2010). It is also clear from studies of both research and data use that these are social processes; practitioners need routines and skilled facilitators to help make sense of evidence and apply it to their work (Coburn & Stein, 2010;
Coburn & Turner, 2011; Honig & Coburn, 2008; Horn, Kane, & Wilson, 2015; Little, 2011; Spillane, 2012). Trusting relationships with researchers or research brokers can also be key to greater research use. They facilitate open flows of information, lead to greater receptivity to research findings, and help tailor research to practitioners’ needs and questions (Coburn, Penuel, & Geil, 2013; Conaway, 2020; Goertz et al., 2013; Hubbard, 2010).

Finally, we know that practitioners’ skills and understandings can make the difference between deriving meaningful insights and implications from research, and not. For example, a study of three districts attempting to use research highlighted that practitioners struggled to “make sense” of research evidence and assess the research claims made by program developers (Coburn, Honig, et al., 2009; Corcoran et al., 2001). Studies about data use have found that practitioners needed data literacy: how to access, manipulate, and analyze data, and “derive instructional implications from it” (Means, Padilla, DeBarger, & Bakia, 2009; Reichardt, 2000; Supovitz & Klein, 2003). Some scholars have suggested a similar need for research literacy. Practitioners may need a basic understanding of study design, sampling, measurement, causation and correlation in order to determine appropriate inferences from research findings (Dynarski, 2008; Penuel, Briggs, et al., 2016).

Others, however, have highlighted how different cultural understandings or assumptions can also maintain a gap between researchers and practitioners. In other words, practitioners may benefit from learning research literacy skills, but they may also have different perceptions of what makes research valid or how conclusive findings should be in order to be usable (Farley-Ripple et al., 2018). For example, Knott and Wildavsky (1980) hypothesized that the likelihood of research findings being incorporated into practice might be related to factors
like an individual’s tolerance for ambiguity. To better articulate what assumptions or understanding matter for research use, we need more studies on what it actually looks like when practitioners use research to inform a decision. Joyce and Cartwright (2020) argue, for example, that effectiveness studies present evidence supporting particular claims about causality or correlation, but when putting that research into “use,” practitioners apply it instead to support entirely different claims about predictions (Joyce & Cartwright, 2020). The research-practice gap persists beyond the level of supply and demand, in this type of micro-process when practitioners interact with and use research.

As SEAs and intermediary organizations are increasingly able to support practitioners with imposed routines, contracted facilitators, and abundant research-based tools, research will make its way onto practitioners’ tables. It will be searchable and simplified, and practitioners will be required to look at it and cite it. But examining what that imposed use looks like in practice reveals the continued significance of—and the challenges posed by—this last element of the research-practice gap.

This study therefore addresses the following research questions:

1. How do district and school practitioners use and interact with their state’s research-use resources and meet their state’s research-use requirements?

2. When and how do practitioners encounter and resolve obstacles to research use?

A Sociocultural Perspective

My investigation of research use’s micro-processes is informed by sociocultural theories on cognition and policy implementation (Levinson et al., 2009; Spillane et al., 2002; Vygotsky,
Weatherley and Lipsky (1977) argued that the real meaning of a policy is found not in the policy itself but in the way implementers put it into practice, and showed that implementers innovated coping mechanisms and made significant accommodations to make policy work for them. Later studies of policy implementation further established that practitioners negotiate new meaning out of policy directives, filtering out some ideas, (mis)understanding some ideas, and co-opting or appropriating others (Levinson et al., 2009; Spillane et al., 2002). They do not do this on their own, but as actors embedded in communities (Coburn, 2001). Therefore, a concept like “evidence” does not have a static, abstract meaning, but rather multiple, context-specific meanings that are constructed by people using evidence in authentic tasks and talking about evidence with colleagues (Hill, 2001; Spillane et al., 2002; Wenger, 1999). The research-practice gap, then, persists in the meanings that are not shared between the two separate communities or “cultural worlds” of research and practice (Coburn et al., 2013; Farley-Ripple et al., 2018; Stein & Coburn, 2010). From a sociocultural perspective, the imposition of instrumental research use relies on the successful transfer of a particular set of meanings—what is research evidence and how is it properly used?—from the SEA to its schools and districts.

To some degree, elaborated tools can help to reify and transfer particular meanings across social boundaries. Functioning as “boundary objects” (Wenger, 1999; Zietsma & Lawrence, 2010), tools capture ideas and make them portable. They thereby serve as a source of common ground or a conduit for shared knowledge between communities, and they can coordinate work across communities. For example, pre-service teachers use rubrics and student work artifacts as boundary objects to build coherence between conversations in their clinical work and their coursework (Nolen, Horn, Ward, & Childers, 2011). Success For All
similarly embedded research-based knowledge into its curricular tools to bring that knowledge directly into classrooms (Datnow & Park, 2010). The SEA’s implementation of the evidence requirements relies heavily on such boundary objects: a pyramid-shaped diagram reifies the hierarchy of ESSA’s four levels of evidence; the WWC’s and Evidence for ESSA’s color-coded ratings reify particular criteria for strong versus weak research evidence; and the SEA’s handbook reifies a sequential understanding of when research should be used.

However, tools are still limited; practitioners interpret and add meaning back into tools as they put them into practice. Reifications are only abstractions of ideas, stripped of contextual information and nuance (Wenger, 1999). As Wenger argues, a procedure can “reify a concept so that its application is automatic,” but it can also “hide broader meanings in blind sequences of operations.” What meaning practitioners make out of tools comes from their participation in its use, and therefore too much reliance on prescriptive tools can impede the kind of constructive, participatory knowledge-building that is necessary for learning. Compliance with a procedure can replace true understanding of the work it has simplified. Wenger calls this the “double edge of reification.” The limitations of tools show up in the literature on data use. For example, data reports reify particular representations of students’ learning and not others; organizing data by student subpopulation can make diagnoses based on subpopulation more automatic and obscure other possible diagnoses (Horn et al., 2015). Even the words in implementation tools do not carry static meanings (Hill, 2001). Thus different assumptions about the term “evidence” or different understandings of decision-making can still layer onto or be held in tension with the meanings reified in the SEA’s tools.
Therefore, to understand imposed research use, I examine how practitioners make sense out of their SEA’s tools and expectations by making them work. To do so, I focus on micro-processes, or patterns of on-the-ground, situated interactions (Little, 2011). Micro-process studies recognize the importance of practice and aim to identify the types of discursive moves or interpretive work that people engage in (Little, 2011; Park, 2018). As school and district practitioners enact imposed research use, I attend to the micro-processes through which they negotiate meaning out of the SEA’s tools and expectations.

Methods

Setting

As noted above, this study took place in a state that has chosen to use ESSA’s revised evidence requirements to push schools’ and districts’ decision-making practice closer to the ideal of instrumental evidence use. The SEA’s implementation approach focused on building local capacity and external oversight for identifying data-based needs, and articulating processes by which local organizations would select interventions based on the strength of their research support and their fit with local needs.

I was interested in how practitioners responded to the SEA’s new expectations for instrumental data- and research-use—particularly how they made use of the SEA’s tools and supports. Informed by my sociocultural framework, I purposively selected sites that reflected a diversity of practice contexts, and then looked for patterns and contrast in the ways participants used data and research in their instances of decision-making.
One dimension of variation was district settings and demographics. Prior studies find that decision-makers look for research evidence from districts or schools in similar urban-rural settings or with similar student demographics to their own (Nelson et al., 2009), meaning that differences in context may shape practitioners’ perceptions of the feasibility or appropriateness of research-based decisions. The three districts in this study reflect a wide range of contexts (see Table 6).

The second dimension of variation was in the tasks and decisions required of each school or district site. Although all school improvement plans needed to meet ESSA’s evidence requirements, the SEA had different processes for schools and districts based on their formal designations. The specific kinds of decisions practitioners needed to make, and the conditions under which they made them, therefore varied based on their designations. Specifically, the SEA aligned its processes for state turnaround (for the overall lowest-performing 3% of schools) and federal comprehensive support and improvement (CSI; for the lowest-performing 5% of Title I schools), meaning that schools in both designations\(^3\) faced similar regulatory pressures, if slightly different timelines and different funding levels. Both turnaround and CSI schools receive support from an external consultant, use the same templates and guidance materials, and will submit improvement plans for review by the SEA. Schools identified for targeted support and improvement (TSI; for schools with low-performing subgroups) have no promised funding, no external consultants, and improvement plans that will be only reviewed by the district. The sites in this study include one turnaround school, one CSI school, and one district with three TSI schools.

\(^3\) Schools were not double-identified meaning that a designation for turnaround supersedes a CSI designation.
Table 6
Research sites and participants

<table>
<thead>
<tr>
<th></th>
<th>LEA A</th>
<th>LEA B</th>
<th>LEA C</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Context</strong></td>
<td>Rural</td>
<td>Urban</td>
<td>Suburban</td>
</tr>
<tr>
<td><strong>Total enrollment</strong></td>
<td>2,893</td>
<td>22,819</td>
<td>6,412</td>
</tr>
<tr>
<td>% economically</td>
<td>100</td>
<td>59</td>
<td>37</td>
</tr>
<tr>
<td>disadvantaged</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% English language</td>
<td>31</td>
<td>21</td>
<td>6</td>
</tr>
<tr>
<td>learners</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% students with</td>
<td>14</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>disabilities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student race/ethnicity</td>
<td>39% White</td>
<td>43% White</td>
<td>69% White</td>
</tr>
<tr>
<td></td>
<td>54% American Indian</td>
<td>38% Hispanic</td>
<td>19% Hispanic</td>
</tr>
<tr>
<td><strong>Schools</strong></td>
<td>12</td>
<td>38</td>
<td>10</td>
</tr>
<tr>
<td>Title I</td>
<td>12</td>
<td>19</td>
<td>3</td>
</tr>
<tr>
<td>In state turnaround</td>
<td>5</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Identified for federal</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>CSI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identified for federal</td>
<td>7</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>TSI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Interview participants</strong></td>
<td>• Arbor principal*</td>
<td>• Blake principal*</td>
<td>• Director of Elementary Teaching &amp; Learning*</td>
</tr>
<tr>
<td></td>
<td>• Arbor teacher leader (3rd grade)*</td>
<td>• CSI consultant*</td>
<td>• Assistant principal, Title I Coordinator*</td>
</tr>
<tr>
<td></td>
<td>• K-12 Schools Director, Curriculum Director</td>
<td>• Focus Schools Director*</td>
<td></td>
</tr>
</tbody>
</table>

* Interview participant was present in observed meetings.

LEA A is small and remote, with five out of the 12 schools designated for state turnaround, making it one of the lowest performing districts in the state. Data collection in this district focused primarily on one school identified for turnaround, Arbor Elementary. LEA B is one of the largest districts in the state, with seven out of its 38 schools either in turnaround or

---

* School names are pseudonyms.
CSI. Data collection in LEA B focused on one of its CSI schools, Blake Elementary. Finally, LEA C is a mid-sized suburban district with no schools in turnaround or CSI, and only three schools identified for TSI.

**Data Collection**

I began data collection in the spring of 2019 by attending two of the SEA’s trainings related to ESSA’s evidence requirements and evidence-based practices for school improvement. I audio-recorded these trainings and collected documents and digital resources provided to district participants.

I then recruited districts from the participants in those SEA trainings, and observed at least one subsequent meeting in each district. I asked participants to identify what meetings would involve discussion of ESSA’s evidence requirements or the selection of evidence-based practices. In all cases, I observed meetings that involved practitioners considering data and/or research, and moving toward at least a preliminary decision about needs or interventions. Observations ranged from 20 minutes to five hours, reflecting differences in the decision-making processes each site was engaged in, with a total of eight hours of observation data.

In LEA A, I observed a full-day, end-of-year staff meeting at Arbor Elementary. The meeting was optional and open to all teachers; there were five teachers, an instructional aide, and the principal present. The aim of the meeting was to get a “jump start” on identifying needs and possible strategies in preparation for their turnaround consultant who was expected to begin sometime during the summer. With the principal as their main facilitator,
the Arbor staff began by using their available assessment data to “look for patterns,” what strategies worked, and “what are we missing?” They finished by discussing “what actions must occur as our next steps.”

In LEA B, I observed two school leadership team meetings at Blake Elementary; the leadership team consisted of the principal and seven teachers, one from each grade level. Prior to my observations, Blake began its CSI process with an onsite needs assessment conducted by their external consultant. Per the SEA’s policy, they were assigned a consultant with whom they had no prior relationship. This consultant synthesized the needs assessment data and conducted a root cause analysis, then presented those findings and analyses to the Blake leadership team. I observed two subsequent meetings: one was the consultant’s final meeting at Blake where she facilitated their process of selecting four focus areas based on the provided data and laid out their next steps for CSI planning. The second observed meeting was facilitated by Blake’s district and university coaches where the leadership team discussed and drafted evidence-based strategies for each of their focus areas.

In LEA C, I observed a district-level meeting of instructional coaches facilitated by the curriculum director who had attended the SEA’s training. The topic was “ESSA Evidence,” and the curriculum director used the time to show her colleagues ESSA’s tiers of evidence and the online research clearinghouses, and to convey her expectation that any future program and curriculum adoptions (regardless of TSI or Title I status) must meet the top three tiers of evidence. With no deadlines approaching for TSI plans, the team instead discussed how the SEA’s tools might inform their upcoming early literacy plan.
All observations were audio-recorded and fieldnotes include partial transcriptions; observations in LEA A were conducted in person, observations in LEA B were conducted via video conference, and LEA C provided a video recording of their meeting. Fieldnotes and transcriptions included notes about pauses, body language, and verbal tone. I also requested relevant documents such as meeting agendas, Powerpoint slides, and draft plans when available.

I followed the observations with semi-structured interviews with at least two individuals in each district, aiming to include at least one district-level and one school-level leader, and prioritizing individuals who were involved in the observed meetings. Interview questions fell into three categories: (1) participants’ overall understanding of the term “evidence-based” and the relevance of “evidence-based practices” in their work; (2) participants’ reflections on their experience in the SEA’s training, what they expected, and what, if anything, they have subsequently thought about or used; and (3) providing context for the observed meetings. Questions in the last category varied somewhat based on participants’ roles and included follow-ups based on specific observed interactions.

**Data Analysis**

Analysis began with close reading and grounded initial coding of all observation fieldnotes and interview transcripts (Charmaz, 2006). Two strategies guided this initial coding. First, given my interest in how participants learn about and enact the new evidence requirements, many of my initial codes for the observations described participants’ actions each time they arrived at (or attempted to arrive at) a decision, such as “seeing growth in numbers,” “interpreting why the numbers are these numbers,” and “using personal example.” Framing
my coding in this manner helped to maintain a focus on *process*—how practitioners use or respond to different forms of evidence to draw conclusions—rather than merely topics.

Second, recognizing how participants’ use of language can capture their particular meaning making and perspectives, I also used *in vivo* codes for both observations and interviews. For example, coding one participant’s repeated use of “own/owning the data” provided a useful way to encapsulate her stance on evidence-based decision-making, while coding another participant’s repeated reference to having or knowing “the answer” highlighted her different perspective.

In a second round of axial coding, I re-read each document and grouped similar codes. I then wrote analytic memos about each of the three LEAs, synthesizing the prevalent codes across the observation, interview, and documentary data. It was at this stage, particularly in connecting observations with interviews, that major themes about gaps and challenges emerged across all three sites. For example, the memo-writing process for LEA A surfaced tensions such as “insider/outsider dynamics regarding evidence,” “the legitimizing and not-legitimizing power of research,” and “when and how to bridge multiple ways of knowing what is working.” As a last stage of analysis, I returned to the literature on research use and sociocultural cognition to put these observed tensions and patterns into theoretical and empirical context.

**Micro-Processes of Making it Work**

This study investigates how school and district practitioners interpreted and implemented their SEA’s tools for imposed instrumental research use. In the end, as required by the SEA, all practitioners had selected or were prepared to select qualifying evidence-based
interventions. And in their processes, they all referenced or interacted with the SEA’s tools and facilitators. Nevertheless, their decision-making practices sometimes took on characteristics of symbolic use or ceremonial compliance. This in itself is not a surprising finding. However, by observing the moments when practitioners slipped away from instrumental use, I identified patterns that reflect the persistent difficulty of bridging the research-practice gap, even while tools and facilitators ensured that research was present.

Symbolic and ceremonial types of use, I argue, were the result of practitioners innovating ways to cope when the meaning of research evidence seemed indeterminate or incommensurate with their own understandings of evidence—but they were still pressured to show that they had “used research.” They deferred research use to a different time and different person rather than integrate it into their discussions, complied with the SEA’s requirements while remaining unpersuaded, and chose to either accept or reject research-based conclusions wholesale rather than blend research insights with their own. I describe each of these micro-processes below.

**Deferral of Research Use**

Though they were in very different situations, practitioners across all three sites chose to delay research use and to delegate it to an expert either in their organization or externally. This was more than just symbolic use; practitioners explained that they wanted to end up with programs and practices that were backed by research. But research use was considered to be a time-specific activity, separate from the kinds of conversations schools needed to have in their decision-making process.
In LEA C, for example, the Title I coordinator attended a district coaches’ meeting about the evidence requirements, research clearinghouses, and tools. In the meeting, the curriculum director showed the team where to find these tools and walked through how to use them. Later, the Title I coordinator reflected on how she imagined her colleagues in the district would use what they had learned:

As far as how much they are going to use it, they probably aren’t going to think a lot about it until they need to, and when they do, then they know who they would call. So you know I think just knowing who your resources are and you know who specializes in this or that is important as well. […] And you know, typically a lot of those purchases have to be approved. And so you know whether it’s on the front end or the back end, if they’re going to, I’m sure hear from [the curriculum director] if they’re buying something that would not be a wise option.

Two points are worth noting in this explanation. First, she conveyed a perception that using the SEA’s research tools was something that only one person needed to “specialize in,” rather than something for everyone to participate in. And second, that research use was something to do only when you “need to,” which could be “on the front end or the back end” of a decision.

At Blake Elementary in LEA B, references to research and “the literature” were technically present throughout the school improvement planning process, thanks to the SEA’s turnaround consultant and district coaches. Blake’s external consultant provided the school with citations for “resources in the literature,” such as John Hattie’s (2012) Visible Learning for Teachers, which she believed were a way to get “good research-based ideas for the practices they need to implement.” Research use was primarily delegated to their consultant and coaches who, as the principal described it, “did the groundwork, presented the evidence to us, and then we were able to include that in our plan.” Yet when presented with these
resources, practitioners tended to use the research information as evidence for and of predetermined decisions.

For example, in the following excerpt, three teachers on the school leadership team worked together to draft a professional learning strategy for the school’s improvement plan. They read aloud from the consultant’s needs assessment report:

Teacher 1: I feel like—and this is an “I feel” statement—they are going to want us to continue with reciprocal teaching practices […]. “[Reading] Content for 2018-19 PD focused on increasing student talk and reciprocal teaching. Although student talk promotes a productive increase in student interactions during learning, student talk does not appear in the literature as a high-leverage instructional strategy for increasing student academic achievement.”

So I feel like that part… is information? […]

“[Continuing to read] Reciprocal teaching, on the other hand, has an impactful effect size of .74 (Hattie, 2012). However, knowledge and skills for full implementation of comprehensive RT, as outlined in the literature for high impact, were not developed or implemented school-wide.”

So that’s something we can use and turn into professional development being differentiated so that it’s able to better develop—

Teacher 2: I don’t see—

Teacher 1: No, no, no. The only reason I’m saying “differentiated” is because our professional learning goal is “differentiated based on needs of instructional staff.”

Teacher 3: Yeah, you could mention “for staff,” right?

Teacher 1: [Writing on template] “Differentiated for staff, based on teacher…”

Teacher 3: For level of understanding.

Teacher 2: [Joking] We may have to take a test.

Teacher 1: Okay, here we go. “PD is differentiated based on teacher needs and level of understanding.” Yep.

77
First, one teacher shared her hunch that their consultant and district coaches would expect them to include reciprocal teaching in the improvement plan in some way. That decision to choose reciprocal teaching due to its high effect sizes was perceived as already made. She then read aloud from the consultant’s report, which highlighted the different effectiveness of increased student talk and reciprocal teaching practices, but there was no subsequent uptake from the group of that research evidence. In fact, the report’s note that increased student talk does not have strong research evidence did not seem relevant or actionable to the group. The group’s continued conversation focused instead on how to fit reciprocal teaching into a professional development goal.

Research evidence only reappeared at the end of Blake’s decision-making process, as they added in research citations to their school improvement plan. As the Blake principal explained,

> We had to look at making sure that any strategies that we’re going to use were supported by research […] and there had to be an explanation of what that research actually was to support the strategies that we were going to be using in the plan. But the conversation was really just looking through all of the needs assessment data and the root causes and trying to come up with ways to address those root causes and developing our plan.

In other words, research was not part of “the conversation” her staff needed to have.

Instead, the principal emphasized the need for conversations to build buy-in: “it doesn’t matter, [chuckles] you know, how effective [it is] and research supports it, if teachers don’t embrace it and own it and implement it the way that it was intended.” The consultant’s tools put research on the table and the SEA’s expectations ensured research’s place in the school’s final plan. Yet its actual role in teachers’ decisions was delegated and deferred: experts used research to recommend particular decisions and to help the school explain their decisions in their final plan. Between those stages of the process, teachers’ conversations focused on how
to make those recommendations *work* for them—how to word their strategy correctly and what type of strategy did they really want to commit to.

Conversations at Arbor Elementary similarly focused on building buy-in without research evidence playing much of a role. The deferral of research use at Arbor occurred at a larger, strategic level because their turnaround consultants were scheduled to arrive months after the staff wanted to begin their planning process. They therefore spent their time moving toward decisions that they knew would later be revised and confirmed with other experts. But it also happened at a smaller, interactional level as the principal repeatedly chose to hold back her research-based knowledge and allow her staff to generate their own understandings of possible interventions.

For example, during the Arbor Elementary planning meeting, the principal worked with a group of three teachers and another administrator looking at their DIBELS data for the past year as they looked for areas of growth and areas that needed attention. The teachers determined that their phonics routines were working because their reading accuracy scores were high, but that they “needed to figure out something more” for fluency. One teacher described how he had used visual-spatial reasoning strategies, which sparked another teacher to think aloud that she could use her magnetic letters manipulatives to help students’ whole word fluency. They wrote down “visual-spatial understanding” on their chart of potential next steps and strategies. In an interview later that day, the principal explained why she did *not* refer to research evidence during that conversation:

> It was really interesting to hear [them] say “Oh I noticed that so-and-so was doing this and you know for my kids they didn’t get here and so I need to…,” you know, some of that talk about automaticity. I mean there’s some definite, like even in my mind, immediately I think of the Tier 2 handbook that goes with the [basal
There are some evidence based practices in there that are specific for automaticity and fluency that we could open up that book and look through there and pick some things. And that’s eventually what will happen along the path. And some things will get thrown out today.

But I think along the path, like language production—we, I don’t have the expertise to help with that piece. We’re going to need to get some help in that area. So I think today is more of getting buy-in, talking through you know where we want to go, what are our target areas, throwing in what we know we already have in school that we’re using or that we need to change and what other things do we need to start looking at and finding resources for.

And then all that will kind of get cemented together in a plan with the consultants. So they’ll knock the rough edges off and bone us in. […] It’s not just ticking the box of turnaround, but I think making sure that we are using evidence-based practices. That there’s not some “great idea” that some teacher has and wants to do [that] gets thrown in [and] we’re having the whole school do that.

Here, the Arbor principal identified “getting buy-in” and “making sure that we are using evidence-based practices” as two separate parts of the school’s decision-making process. To get buy-in, she let teachers “throw out” their own ideas, and chose not to bring up the research-based practices in their curriculum. This was partly possible because she expected the turnaround consultants to be able to provide even more research expertise and a checkpoint on ideas generated by the staff. But it also, like at Blake Elementary, reflected a sense that research evidence is useful in some but not all decision-making conversations.

On the surface, this deferred research use sounds like symbolic use—attention to research came after the emergence of a solution (Coburn, Touré, et al., 2009). Conventional accounts of symbolic use suggest that practitioners engage in symbolic use to justify pre-existing preferences, or out of a lack of time and resources to search for and incorporate research more meaningfully (Coburn, 2010; Coburn, Honig, et al., 2009). An additional reason to delay research use, suggested in these findings, is that teachers and district staff had incremental, multi-stage understandings of the decision process—but time-specific
understandings of research use. Research evidence simply made more sense in later stages of a decision, used to confirm a decision rather than to originate or persuade others of a decision, and could be done in one moment rather than over time. By contrast, local data and practitioners’ perceptions were where solutions and strategies emerged from and were worked out together.

Skeptical Compliance

Sometimes practitioners put the SEA’s research tools into use exactly as they were trained, yet they separated compliance with the requirements and processes from actual persuasion or enlightenment through using research. This was the case in LEA C, where the curriculum and instruction staff prepared to adopt a new literacy program. The SEA’s different tools prioritized and clarified some aspects of research use, but left others opaque to practitioners.

After attending the SEA’s ESSA training meeting, the curriculum director in LEA C convened a meeting with eleven of her district colleagues to share the SEA’s tools and how to use them. She projected the SEA’s infographic depicting ESSA’s four tiers of evidence as tiers of a pyramid along with their definitions. She said:

“So what this evidence-based piece that ESSA provides is a little greater flexibility, but it does increase your responsibility to make sure that the evidence meets the criteria. […] So this slide actually puts all of that evidence piece to practice. So if there’s any slide that really kind of guides us, it’s this triangle. In ESSA […] whenever you use an intervention or you use any kind of strategy, you have to use the top three tiers of evidence in ESSA. Either strong evidence, moderate evidence, or promising evidence when you choose a strategy or a program. Okay?”

Later, to demonstrate how to use the Evidence for ESSA website, she looked up a reading intervention the group was considering for their upcoming literacy plan. Pointing to the results page, she said, “Here you go, you’ve got an ESSA rating that it has strong evidence—
remember, if you go back to this [returns to infographic]: strong, moderate, or promising evidence.” She also modeled how she had used the WWC’s practice guide on out-of-school time strategies to write a grant. She pointed to the WWC’s ratings, “So interestingly, [referring to the screen] there’s minimal evidence on aligning out-of-school time program academically with the school day […] but there’s moderate evidence on adapting instruction individual and small group needs.” As colleagues nodded along, the curriculum director reinforced the idea that research evidence could be evaluated according to tiers, and that those evaluations or ratings directly informed what were or were not acceptable decisions. The SEA’s pyramid diagram and the aligned research clearinghouses thus worked as boundary objects, transferring these particular understandings of instrumental research use clearly and directly into LEA C’s work.

Nevertheless, the Title I director in LEA C later explained that she struggled with the inferential leap between finding a research-based program, and knowing whether it will work with their students in her context. She noted that she still had unanswered questions after the curriculum director’s meeting. In her explanation, she separated research use into two distinct efforts: the relatively straightforward task of fulfilling the evidence requirements, which involved understanding the acceptable tiers of evidence and selecting interventions that match them, and the more complex task of looking at research evidence of effectiveness and inferring its implications for actual practice. She explained that when she looks at a research-based intervention, she “internally question[s]” the population of students, “in what ways” is the intervention effective, how updated the research is, whether the research is biased, and whether the intervention is affordable.

Those are always questions that kind of come to mind and I don’t know that I would ever know the answer to that. But I feel like, you know, from my compliance
standpoint, you know, I could at least say, you know, we are utilizing our Title I money in a way that is, you know, we’re using these programs and these strategies which are, you know, under the guidelines of ESSA, are approved. Or at least supported. So I think from a compliance perspective it probably makes me feel better. But from an actual like, you know, usage perspective I think I still have some questions. […] I can justify spending money on this because, you know, it fits within the requirements of ESSA. From an actual student growth perspective, you know, there I think there’s more questions to be answered that I don’t think you know. […] You have to be really careful that you’re not, you know, that you’re making sure that all of that [Title I] money is spent in a way that you know “this is going to impact student achievement [and] how,” you know, and that you can answer that question.

Here, the Title I director highlighted additional understandings of research use that are not embedded in the SEA’s tools. In her work, she felt the need to be able to say, “this is going to impact student achievement” and she felt accountable for having the “answer.” This relies both on an understanding evidence as a means for prediction rather than evidence of effectiveness elsewhere, and on an understanding of evidence as highly conclusive. She listed many important questions about generalizability and worried that “I don’t know that I would ever know the answer” and “there’s more questions to be answered that I don’t think you know.” Some questions could potentially be answered with more information about the research studies; but her sense that some of her questions were essentially unknowable points to an inherent gap in using research on a program’s effectiveness elsewhere to infer its potential effectiveness in practice locally. The Title I director also struggled to match ESSA’s acceptable tiers of evidence with her own high internal bar for confidence in her decisions. This suggests that concepts of confidence and certainty in evidence may also differ for practitioners and researchers. In the end, she was able to use the SEA’s tools to comply with ESSA’s evidence requirements, but not to convince herself that her decisions were necessarily right.
As noted earlier, prior studies of the research-practice gap have highlighted the importance of research literacy skills, and suggested that practitioners may need help knowing how to judge appropriate inferences based on research evidence. Tools and reifications like the pyramid diagram of ESSA’s tiers of evidence and clearinghouses with easily-identifiable ratings can help simplify such inferences. Yet these findings suggest that the gap may go beyond skills into how practitioners think about the nature of evidence. Practitioners face demands to use funding well, raise student achievement, justify their actions, and lead implementation. They bring local knowledge of their context and a “usage perspective” to their decision-making. To engage in instrumental research use, then, they need to be able to reconcile their expectations about research evidence with what answers it can actually provide. Otherwise, as described above, their use of research may simply be ceremonial compliance with the regulations.

**Overriding Other Forms of Knowledge**

Under pressure to meet the evidence requirements, practitioners sometimes leapt to over-simplified conclusions from the available research evidence. This was particularly the case when research evidence contradicted pre-existing understandings about best practices and good curricula. Practitioners tended to let one conclusion override the other—either using research to delegitimize a prior decision, or using other evidence delegitimize the research findings. Practitioners took research evidence to be a very strong signal, even when that evidence was as little as a rating (or even lack of a rating) in one of the SEA’s tools. But they also seemed to struggle to reconcile or merge that signal with other forms of knowledge or understandings of practice.
At the SEA’s ESSA training, practitioners were given time to look up the research ratings for programs that were currently used in their schools and districts. Multiple training participants noted that their programs were missing from the research base, had weak evidence ratings, or small effect sizes. A few used this new evidence to revise their thinking about and implementation of their old program. But others, facing this apparent contradiction, interpreted the research evidence as wholly invalidating their district’s decision or wholly inapplicable. Take for example the contrasting reactions from the Blake Elementary principal and her district coach who found out together that their reading curriculum lacked significant positive results. As she later shared, the principal at Blake Elementary interpreted the poor research evidence as the district’s error:

Principal: We even looked up Futures, which is the curriculum that our district uses for language arts instruction. And that was kind of interesting because it didn’t have really high scores [chuckles] or very, it didn’t really have that much support for it. Interesting.

Researcher: What do you do when you find that out?

Principal: I know, and you’re using it right? Yeah, it’s not something you want to let your faculty know [chuckles], because the expectation is that they are implementing with fidelity so that we all have a common curriculum. And it’s supposed to be aligned with state standards, so. But as far as the effectiveness of it, it wasn’t getting good remarks [sic].

Researcher: That’s tricky.

Principal: That’s interesting. Yeah, because they do put a committee together at the district level to research all that curriculum and come up with the top recommendations. So.

---

5 Although both the What Works Clearinghouse and Evidence for ESSA interpret evidence strength for users (i.e., with color-coded labels such as “Strong” or “Positive or Potentially Positive”), I observed participants making their own interpretation of effect sizes. Evidence for ESSA’s “Frequently Asked Questions” explains that an effect size of +0.10 is small, but one training participant found positive average effect sizes of 0.36, 0.17, and 0.20 for her district’s reading interventions and called them small and “not very good.”

6 Pseudonym.
Somehow, the district committee that was supposed to “research all that curriculum” chose a program that did not have high scores. She concluded that this was something that her faculty should not be told—an error or a problem that would undermine their buy-in and the fidelity of their implementation. Her repeated note that this was an “interesting” finding was diplomatic, but also suggested she was not completely sure what to make of it.

On the other hand, the district coach looked for alternative evidence that would re-confirm that the Futures curriculum and their math curriculum were still good choices. After the training session, she asked her more senior colleagues to explain the adoption process.

And they at the district level, they had really great answers. They were like, “it incorporates […] constructivist mathematics combined with this, and then we have this for formative assessment, and we have this to scaffold and build background knowledge and build vocabulary in it. And it has these and it spirals in this way.” I mean, they knew the answer.

But the principal didn’t, wasn’t able to articulate that. But that is one of my principals of a more underperforming school, so I don’t know if that was just particular to that principal.

Whereas the principal accepted the clearinghouses’ implication that her district had made a poor decision, the coach rejected it. She was satisfied that the district had “the answer,” even though that required adopting a different interpretation of the evidence requirements: the programs may have lacked ESSA’s level of research evidence, but the embedded practices were still generally supported by research concepts. The clearinghouses and tiers of evidence reified clear, but simplified meanings of evidence. Practitioners, searching for “the answer,” used that evidence to support similarly simplified claims about whether a program was good or bad.
The teachers at Blake also struggled to hold contradictory evidence in tension. As part of their CSI process, they received a needs assessment report conducted by their external consultant. This needs assessment was a form of internal research, combining classroom observations, surveys, document reviews, and focus groups. The SEA’s expectation was that the school would use this research evidence to decide on their areas of need, but doing so expected teachers to reconcile their own observations and perceptions with the consultant’s research findings. In practice, teachers were pressured to accept the consultant’s findings over their own, and they responded by rejecting them.

For example, in one school planning meeting, the consultant addressed one of the areas of need she had identified in her report—that the Blake teachers were not consistently teaching their students routines for positive behavior. She said, “And while I heard very clearly from the leadership team that, ‘hey, maybe your team didn’t see the routines in the classroom’ […] there are other sources of evidence, right, that show us that behavior is a challenge for you all here.” As they returned to their planning work, the consultant announced, “So before we dive in, let’s set our task. [Reading aloud from the Powerpoint in chorus] With an open mind, read the suggested root causes and data collected as evidence of the root cause.” In these opening remarks, the consultant defended her team’s observational data over the Blake team’s argument that they actually were teaching those routines and suggested that the teachers needed to be more “open-minded” about the research findings.

In the next Blake leadership team meeting, which the consultant did not attend, one group of teachers read through the report and ruled out another potential focus area on the same
grounds. After reading aloud one assessment which described the percentage of lessons that were observed engaging in a particular practice, two teachers commented:

Teacher 4: So I think we should not focus on this one, simply because—

Teacher 5, jumping in: And all I can think about is the fact, at the time of the year they came, [inaudible] could not be helped, but what I was doing, the one time they observed me for ten minutes, was small structured intervention, and we were already halfway through the intervention by the time they showed up and by the time we rotated, they had left. So it was like, no, I didn’t communicate that, but it was also targeted [intervention time].

Teacher 4: Yeah.

Teacher 5: And then we had other things going on so they never made it back to my class. So it’s hard, you know?

As one teacher said, “I feel like this is already happening” even if the consultants “didn’t necessarily see it.” Convinced that their anecdotal and personal knowledge about school-wide practice was right, and the consultant’s team had failed to observe it, this team of teachers chose not to address this area of practice at all. Rather than discussing how both sources of evidence might be true and inform their thinking, the staff and consultant at Blake engaged in a back-and-forth over whose evidence was right.

The SEA’s tools reified particularly simple understandings of research evidence; programs either qualified or did not qualify under ESSA’s evidence criteria. The SEA also contracted consultants to conduct internal research reports at their low-performing schools, and these reports also summarized schools’ needs into short descriptions and percentages. First, these findings show practitioners reading and engaging with these forms of research evidence. But importantly, these findings also show that practitioners struggled to make sense of research evidence together with contradictory decisions or perceptions. These forms of evidence were incommensurable. Can a program be ineffective and use best practices? Can I know that we
are teaching routines and researchers find that we are not? Instead, practitioners came to their own conclusions, letting one form of knowledge override others.

Summary

The SEA in this study attempted to impose instrumental research use for its schools and districts, relying on tools and requirements to extend their reach into practitioners’ decision-making work. These implementation tools served as boundary objects, carrying certain understandings from the research community into the practice community—namely, a hierarchy of evidence strength, criteria for qualifying evidence, and the need to use evidence to justify decisions. Practitioners interacted with these research-based tools. They learned the tiers of evidence, searched for evidence on possible programs, read pieces of research evidence, and cited research in their improvement plans. But while doing so, they also withheld their skepticism about whether research really could provide the evidentiary confidence they needed; separated their conversations into research-free development of solutions and research-based confirmations of those solutions; and felt compelled to either rely on research or their own knowledge but not both.

Each of these patterns in practice highlight the persistence of a research-practice gap. I argue that each micro-process reflects practitioners’ efforts to “make it work”—that is, to still use the tools and meet the expectations, despite this gap. In Figure 2, I show how each micro-process responds to gaps between research evidence and practitioners’ understandings of decision-making and knowledge. In the top row, practitioners were able to connect research and practice by deferring research use; by beginning instead with local knowledge and conversations among colleagues, they generated solutions and buy-in, which could then be
confirmed with research. In the second row, practitioners were not able to connect research and practice; putting the SEA’s tools into use meant transferring evidence that a program had worked elsewhere to support claims that the program would work here. At least for one practitioner, this was an insurmountable inferential leap, resulting in skeptical compliance. Finally, in the third row, practitioners found research evidence to be incommensurable with their own understandings; when these sources of knowledge contradicted, they let one override other forms of knowledge, rather than merge insights from both.

Figure 2
Summary of Micro-Processes and the Research-Practice Gap

Discussion and Conclusion
As imposed research use has gained traction as a reform movement, researchers have developed an increasingly comprehensive literature on the capacities, tools, routines, and other factors that impede or enable substantive connections between research and practice.
This SEA’s approach to ESSA’s evidence requirements appeared to respond to this literature by focusing on capacity-building and the provision of several highly-elaborated tools. Yet the findings presented here suggest that while this approach may shift practice, research use still often happens in symbolic and ceremonial ways.

I argue that a sociocultural perspective helps us understand why the imposition of instrumental research use fails to fully bridge the research-practice gap. Past articulations of the gap focused on practitioners’ inattention to research (cf. Weiss et al., 2008) or the lack of research dissemination (Kaestle, 1993; Knott & Wildavsky, 1980). We have increasingly become aware that practitioners do pay attention and have access to research; what matters is what they do with it (Farley-Ripple et al., 2018; Stein & Coburn, 2010). In this study in particular, I found that practitioners read, used, and talked about the research tools and resources provided by the SEA. But the SEA’s attempts at greater prescriptiveness, or greater reification of the meaning of evidence and its use, were still frustrated by the ways practitioners interpreted and applied their tools. In other words, a gap remained as practitioners negotiated meaning through participation with those tools (Wenger, 1999).

These findings contribute to our conceptualization of the research-practice gap and how practitioners’ understandings of evidence and decision-making butted up against those embedded in their research tools. They also have implications for the kinds of training or support that might better bridge that gap.

First, though the SEA’s tools encouraged consulting research at an early stage to identify possible interventions, practitioners recognized many stages and strands of the decision-
making process, and research was only relevant for confirming their choices at the end. This could partly be the result of capacity limitations, since delayed or symbolic use has been associated with time and resource constraints (Coburn, 2010). Practitioners believed that not everyone could be experts in research use, so it may have made more sense to delegate and defer research use to a final check rather than engage in research use on an ongoing basis. And it could also partly be due to the well-documented tendency for practitioners to prefer and be more persuaded by their own experiences or local, relevant evidence (Birkeland et al., 2005; Corcoran et al., 2001; Nelson et al., 2009). But practitioners did not always prefer local evidence over research evidence. Rather, practitioners understood “evidence” as serving many different purposes during a decision, from generating ideas, to building consensus around an idea, to approving or certifying an idea. Practitioners were able to make research use work for them when they separately used local data and collegial expertise for some purposes, and research for others. One practical implication would be for implementation tools to highlight other ways to use research, not just to gauge effective interventions, but also to expand conceptual understandings (Farrell & Coburn, 2016). Even when practitioners are required to execute their school improvement decisions on a particular timeline, decisions are really made incrementally and opportunistically (M. D. Cohen, March, & Olsen, 1972; Weiss, 1980). Therefore efforts to imposed research use cannot target specific decision events and expect to truly influence practitioners’ decision-making.

Second, even when using research to confirm a decision, practitioners distinguished between ESSA’s criteria for strong, moderate, or promising evidence, and their own criteria for evidence. ESSA’s tiers are based on research design and an understanding that well-designed, well-implemented studies can offer different levels of confidence in their conclusions about
an intervention’s effectiveness. Traditionally, the research use literature has suggested that practitioners might be better able to use research if they had more training in this type of research literacy (e.g., Dynarski, 2008; Penuel, Briggs, et al., 2016). From a practitioner standpoint, however, confidence came from being able to say “this will work here.” Transitioning from one understanding of evidentiary certainty to another may involve more than just knowledge about generalizability or external validity. Researchers need to understand the pressures practitioners face and the questions they need to answer while making decisions, and practitioners need to understand what prior research can and cannot answer. Bridging this gap might also involve helping practitioners build new research evidence that “this is working here” (Joyce & Cartwright, 2020; Penuel & Farrell, 2016).

Finally, practitioners struggled to weigh and consider research evidence alongside their own professional and local knowledge. Ideally, research use would add new insights from research to practitioners’ overall working knowledge and understanding of their problems and possible solutions. Instead, practitioners either accepted research-based conclusions and delegitimized alternative conclusions, or did the reverse. This suggests that perhaps part of bridging the research-practice gap is finding ways for those separate bodies of knowledge to be more commensurable—valued equally, articulated in similar language and similar detail. Research-based tools should not be so simplified that they prevent practitioners from identifying insights from research. For example, an emphasis on ratings may have contributed to practitioners’ sense that a program either was good or bad. Emphasizing lessons about implementation or program components could help practitioners connect research evidence to their own knowledge about local implementation and best practices.
These findings suggest a need for greater attention to the cognitive micro-processes involved in research use, and particularly in the ways practitioners reconcile or revise what they know about their practice with potentially contradictory research. This study was limited in scope; therefore further research could examine whether these patterns and micro-processes exist in other contexts, and what practitioners’ research use looks like on a longer time frame. Studies in other states could also explore how different implementation tools affect practitioners’ decision-making.

Taken together, this study’s findings provide mixed hope for imposing instrumental research use in schools and districts. On one hand, we see immediate uptake of user-friendly tools and an earnest effort by many practitioners to make meaning out of the research that was available to them. But on the other hand, it is clear that current tools and training were not yet sufficient for practitioners to make the kind of inferences and strike the kind of balance needed to use both their practical knowledge and research at the same time. The research-practice gap persists beyond the provision of tools and facilitators, due to different understandings about the nature of evidence and what makes it usable for decisions. Instrumental research use thus remains elusive to imposition.
Conclusion

The Every Student Succeeds Act presents new opportunities to promote evidence-based decision-making. Its definition of evidence is more specific and realistic in comparison to its predecessor in the No Child Left Behind Act. Additionally, in the intervening years, new supporting clearinghouses, intermediary organizations, and research-based resources have proliferated in the field of education. The release of ESSA’s evidence requirements therefore prompted a great deal of curiosity and hope for how they might bring about a new era of evidence-based decision-making (e.g., Results for America & Chiefs for Change, 2018; Dynarski, 2015; Penuel & Farrell, 2016).

These two studies of the implementation of ESSA’s evidence requirements suggest reason for both optimism and concern. In the first paper, I find that many SEAs envisioned a new, evidence-based mindset for local decision-makers, whether that meant using new tools to be “informed consumers” of research and to evaluate and select interventions, or supporting efforts to systematically build up and disseminate an evidence base on local practices. I also observed that half of the SEAs in the sample had found external research findings relevant and useful to their work and expected that their schools and districts would similarly be able to directly apply interventions supported by external research to their improvement efforts. What was surprising were the four SEAs who supported the idea of evidence-based decision-making but eschewed an instrumental approach in favor of a loose, conceptual form of sanctioning use. These SEAs represented vastly different degrees of research capacity and authority, yet all hoped to empower local practitioners to continue their own context-specific or innovative practices by using ESSA’s fourth tier of evidence.
In the second paper I find that school- and district-level decision-making in one of the instrumental approach states was clearly impacted by the SEA’s tools, assistance, and pressure. In a high-performing district, staff used the clearinghouses and guidance on ESSA’s tiers to evaluate potential new literacy programs and to identify weaknesses in their existing reading program. Staff at a low-performing school complied with their improvement consultant and coaches by selecting research-based strategies and including research citations in their plan. Yet using the SEA’s tools and meeting its expectations was not straightforward. Practitioners found ways to cope when research evidence was contradictory with their prior knowledge, uncertain, or unpersuasive.

Looking across both studies, a key question is whether the imposition of evidence use assists or hinders the advancement of substantive evidence-based decision-making:

First, as Weiss et al. (2008) point out in their study of the Safe and Drug-Free Schools program, policies that include evidence requirements ensure that research is “on the table.” All eight SEAs in the first study used research-based tools to steer their schools and districts toward effective practices and programs, and all three districts in the second study had research available and present while they discussed their needs and potential interventions. Moreover, for administrators in the SEAs adopting instrumental approaches, the clarity of the requirements were helpful for pushing their schools and districts toward more rigorous sources of evidence. Staff in one of those districts agreed, it was very clear to them how to select programs with strong levels of research evidence.
However, the first paper makes clear that at least several SEA administrators aimed to promote evidence-based decision-making without the pressure of the evidence requirements. That is, while ESSA effectively made evidence a priority for the SEAs, when they adapted it to their own understandings of utility and relevance, these administrators felt that the specific tiers might actually impede their improvement efforts. Not wanting the requirements to discount what they saw as promising local practices, they offered their schools and districts broad research-based frameworks to work within, rather than evidence ratings and research clearinghouses. Additionally, recognizing that local capacity for evaluation is still nascent, they encouraged even less-rigorous forms of evidence gathering in order to build an evidence-based way of thinking. Done well, this approach has the potential to bring the worlds of practice and research closer together, as many in the field hoped ESSA would (Dynarski, 2015; Farley-Ripple et al., 2018; Kane, 2017; Penuel & Farrell, 2016; West, 2016). The irony is that this approach also involved the most ceremonial or limited compliance with the requirements.

Similarly, the second paper highlights how the very process of pushing school and district leaders’ decision-making practice closer to an imposed model of instrumental use made it more difficult. Some practitioners tried to make instrumental research-based decisions, as they were pressed to, but had to override their own practical knowledge or suppress reasonable questions about validity and generalizability to do so.
### Appendix A

<table>
<thead>
<tr>
<th>Approach and Prior Operationalization</th>
<th>Potential Indicators in ESSA Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Instrumental</strong></td>
<td></td>
</tr>
<tr>
<td>- Decision-makers use evaluation results as the basis for decisions</td>
<td>- School improvement plans need to include analyses of needs assessments and evidence bases justifying intervention/strategy choices</td>
</tr>
<tr>
<td>- Decision-makers consulted an evaluator/expert</td>
<td>- SEA approval process includes criteria for rigor of evidence-based decision</td>
</tr>
<tr>
<td>- Decision-makers cite a particular study or specific finding (Weiss et al., 2005)</td>
<td>- SEA provides technical assistance and other supports (perhaps through external partners) that include interpreting the evidence base for possible interventions/strategies</td>
</tr>
<tr>
<td><strong>Conceptual</strong></td>
<td></td>
</tr>
<tr>
<td>- Decision-makers read or talk in-depth about research literature</td>
<td>- School improvement plans invoke research-based ideas without specific citations</td>
</tr>
<tr>
<td>- Language, conceptual categories, or ideas from the research literature plays a role in how problems are defined or particular solutions are seen as appropriate, usually at a later date and sometimes without invoking the research explicitly (Coburn, 2010, p. 173)</td>
<td>- SEA approval process confirms research basis for general assertions about the research</td>
</tr>
<tr>
<td>- School improvement plans invoke research-based ideas without specific citations</td>
<td>- SEA provides technical assistance that involves reading/discussing research literature</td>
</tr>
<tr>
<td>- SEA approval process confirms research basis for general assertions about the research</td>
<td></td>
</tr>
<tr>
<td><strong>Symbolic</strong></td>
<td></td>
</tr>
<tr>
<td>- Attention to research comes after the emergence of a solution,</td>
<td>- School improvement plans affirm use of evidence without details, explanations, or citations (e.g., a checkbox rather than a narrative)</td>
</tr>
<tr>
<td>- Evidence of selective use of research,</td>
<td>- SEA approval process does not include criteria checking the evidence base for selected interventions/strategies</td>
</tr>
<tr>
<td>- Participants evoke evidence in very general terms (e.g., “research says”) to generate legitimacy for a particular solution (Coburn, 2010, p. 172)</td>
<td>- SEA provides minimal technical assistance or support regarding evidence-use and/or guidance allows districts to search for confirmatory evidence for pre-determined interventions/strategies</td>
</tr>
</tbody>
</table>
Sanctioning

- Decision-makers know very little about relevant evidence
- Decision-makers question the validity or applicability of research, but choose the research-based program anyway
- Decision-makers select from an official, approved list (Weiss et al., 2005)

- School improvement plans select interventions/strategies from a SEA-provided evidence-based list (produced by SEA or external sources) without engaging the research bases behind those interventions/strategies
- SEA approval process confirms that decisions are backed by research-based lists
- SEA provides technical assistance or other support that simplifies the selection process for possible interventions (e.g., moves directly from identifying a need to selecting a matching intervention)
References


