Data for Continuous Programmatic Improvement

This book addresses the issue of data use in educator preparation programs toward continuous programmatic improvement. With an aim to increase the rigor in both research and practice in educational administration and teacher education, this volume will analyze the longstanding quality concerns about teacher and leadership preparation and standards for programs and educators, as well as controversies concerning national accreditation and federal efforts to mandate program reporting data. By exploring the policies and practices that influence departments of education, this volume examines the increasing pressures to improve institutional functioning within a complex system of university, state, and national structures and organizations.

Ellen B. Mandinach is a senior research scientist and the director of the Data for Decisions Initiative at WestEd, USA.

Edith S. Gummer is the executive director for data strategy at the Mary Lou Fulton Teachers College at Arizona State University, USA.
This book is dedicated to all of the educators who understand that data come in multiple forms, who believe that evidence is important to decision making, and who realize that improvement is continuous work. Your striving to prepare your students, whether pre-K–12 or teacher candidates, is what will ultimately results in better educators.

Ellen dedicates this book to Eli and Houdi, who make every day a lesson in humor. You provide a better understanding of collecting all sorts of data to inform continuous improvement outside the realm of education. You have made the geographical transition much easier by pulling me away from the computer periodically or lying across it (Houdi) so I take time away from writing.

Edith dedicates this book to the next generation of the family, Kayla, Wren, and Jackson. Pictures and video chats are my inspiration in continuing this work.
Contents

List of Figures ix
List of Contributors x
Acknowledgments xi

PART I
Introductory Chapters 1

1 Introduction and Overview 3
   ELLEN B. MANDINACH AND EDITH S. GUMMER

2 Landscape View of the Continuous Improvement Process 16
   EDITH S. GUMMER AND ELLEN B. MANDINACH

PART II
From the Researchers 33

3 Creating a Data Culture in Educator Preparation: The Role of the States 35
   MICHAEL ALLEN AND CHARLES COBLE

4 Building Capacity and Commitment for Data Use in Teacher Education Programs 68
   CHARLES A. PECK AND SUSANNAH DAVIS

PART III
From the Field 93

5 Moving Toward Common Measures to Accelerate Improvement: A Roadmap for Educator Preparation Programs 95
   TRACEY WEINSTEIN AND CHARIS ANDERSON
Contents

6 Connected for Improvement: The Teacher Preparation Data Model and TPP Dashboard 118
ANA QUINTANA, BRYAN RICHARDSON, AND CARI REDDICK

7 Learning to Walk the Walk of Continuous Improvement 142
KAYCEE A. SALMACIA AND LIAM HONIGSBERG

8 Visualizing Data: Necessary but Not Sufficient for Educator Preparation Program Continuous Improvement 170
EDITH S. GUMMER

9 Lessons Learned While Building a Data Use Culture in Teacher Education 188
D. SCOTT RIDLEY AND DOUG HAMMAN

PART IV
Conclusions 199

10 Challenges, Opportunities, and Next Steps 201
ELLEN B. MANDINACH AND EDITH S. GUMMER

Index 212
Figures

2.1 Two revolutions for change in education 26
2.2 ASU innovation in teacher preparation 28
4.1 A holistic view of data use practice 72
6.1 Relationships among TPPs, states, and districts 120
6.2 Ed-Fi data model with TPDM extensions 130
6.3 Representation of the TPDM 132
6.4 TPP dashboard connected to Ed-Fi TPDM ODS 139
7.1 Relay Graduate School of Education instructional elements 144
7.2 Relay GSE pathway for measuring academic achievement 145
7.3 Relay GSE teacher alumni perceptions about their Relay GSE experience 167
8.1 Administrator view of site coordinator assessments 175
8.2 Drilling down to the site coordinator level to see candidates at risk 176
8.3 Administrator view of individual teacher candidate 177
8.4 Administrator view of the information reported as a teacher candidate concern 177
8.5 Administrator view of alerts from the iTeachAZ system 178
8.6 Site coordinator view of cohort of teacher candidates 179
8.7 Teacher candidate view of individual performances on assessments 180
8.8 Teacher candidate view of performance assessment 181
8.9 Teacher candidate view of performance relative to cohort average 181
8.10 Feedback provided to the teacher candidate from site coordinator observation 182
8.11 Teacher candidate view of walk-through information 183
8.12 Teacher candidate view of mentor teacher progress report 184
8.13 Teacher candidate view of mentor teacher feedback 184
Contributors

Michael Allen, Teacher Preparation Analytics
Charis Anderson, Deans for Impact
Charles Coble, Teacher Preparation Analytics
Susannah Davis, Oregon State University
Edith S. Gummer, Arizona State University
Doug Hamman, Texas Tech University
Liam Honigsberg, Relay Graduate School of Education
Ellen B. Mandinach, WestEd
Charles A. Peck, University of Washington
Ana Quintana, UPD Consulting
Cari Reddick, UPD Consulting
Bryan Richardson, Bill & Melinda Gates Foundation
D. Scott Ridley, Texas Tech University
Kaycee A. Salmacia, Relay Graduate School of Education
Tracey Weinstein, Deans for Impact
Many people and institutions have influenced our decision to write this book and to collaborate with leaders in the field of policy, practice, and research. Educator preparation programs (EPPs) are being forced by accountability requirements to improve. More important, through institution introspection, the institutions are coming to the realization that they must implement improvement processes for their own edification. Both of us have traveled the country, learning from colleagues and observing what is happening in a number of institutions. These interactions have helped us to understand the landscape in which EPPs must function and recognize the complexities of the systemic nature of the process. We have learned from every interaction and conversation. These activities led us to the conclusion that the collective knowledge of the authors represented in this volume have important messages to the field at large. EPPs are trying to figure out how to proceed in the continuous improvement process. Some are ahead of the learning curve; others are struggling. All institutions meet significant challenges because of the complexity of the enterprise.

We want to thank all institutions and administrators with whom we have come in contact. They have helped us to clarify our thinking. Reading seminal reports from scholars like Charles A. Peck, Michael Allen, and Charles Coble has helped us to understand the policy landscape. Attending meetings like those convened by the Data Quality Campaign have introduced us to how others are thinking deeply about these issues. Interacting with some of the Deans for Impact Institutions such as the Relay Graduate School of Education, Western Oregon University, the University of Texas Rio Grande Valley, and the University of Southern California have brought to the fore issues and challenges that even forward-thinking institutions are facing.

A significant opportunity presented itself when Edith Gummer moved to the Mary Lou Fulton Teacher College at Arizona State University in the role of executive director of the Office of Data Strategy. Edith is living this change process on a daily basis to implement a data culture in the EPP. She is no longer a researcher looking objectively at the field; she is a fundamental change agent. Edith thanks her ASU colleagues for the reality check.
Ellen Mandinach is grateful for the opportunity to have conducted workshops and given a keynote address for the Council for the Accreditation of Educator Preparation, where she has learned a great deal from the attendees. She has also benefited from participating on numerous advisory or working groups for the Data Quality Campaign, where there have been chances to learn from many smart and experienced people from diverse professional organizations.

Both of us would like to acknowledge our home institutions that have provided the opportunity for us to pursue our passion of understanding data use in various venues; WestEd for Ellen and Arizona State University for Edith. Finally, we would like to thank Routledge for recognizing the need for a book like this. We hope it has a positive impact on the fields of policy, practice, and research.
Part I

Introductory Chapters
1 Introduction and Overview

Ellen B. Mandinach and Edith S. Gummer

Purpose of the Book

The purpose of this volume is to contribute to the knowledge base that colleges of education—in particular, educator preparation programs—can use to support their efforts to incorporate the use of data to inform their practice, stimulating the development of data cultures where evidence is used for programmatic continuous improvement, a topic that many institutions already acknowledge as a growing emphasis in the field. This book will help administrators and faculty understand the importance of data for continuous programmatic improvement and the need for and the complexities of creating data cultures in their institutions. Because of the growing pressure toward improvement, it is no longer possible to eschew the use of data. Data use has become a necessity, but it should be viewed as a constructive tool toward institutional improvement. The book acknowledges the challenges. But it explores the opportunities afforded by data use for programmatic improvement. The volume provides examples from leading programs that are effectively establishing data cultures for continuous programmatic improvement. The book provides both a reality check but also a call to action for effective data use. Education must now become more responsive to stakeholders and address the public and policy pressures for improved performance. The use of data is a crucial support of that responsibility.

Emerging Need for Data Use in Educator Preparation Programs

The book addresses the issue of how educator preparation programs (EPPs) must and should use data to inform their practice and for the purpose of programmatic continuous improvement. The trend in education has become very clear of late. The push has been for education to become evidence-based, like other professions. This has been the case for K–12 teachers, schools, and districts. It is also now true for the institutions that prepare educators for their professional careers. Education, as a profession, must become data-driven and evidence-based, just like medicine, business, and
even professional sports. In part, this is why the Institute for Education Sciences in the United States Department of Education was created, to increase the rigor in both research and practice in education.

Education can use models of data use in other professions to develop processes and structures. In medicine, not only must professionals be armed with data to inform their decision making, but the institutions now are using data to improve their performance. Doctors use data to make diagnoses and form a course of treatment much like what teachers do in their instructional practice. Hospitals routinely use data to determine how well integrated and effective services are coordinated. A visit to an emergency room generates an automatic survey about the performance of the staff and the institution. In business, sales figures serve as data to inform future inventory control and purchases. Restaurants, car dealers, and service providers routinely send a customer satisfaction survey upon completion of a meal, purchase, or service. Sports are increasingly becoming data-driven. Even before the book and movie *Moneyball* (Lewis, 2004) made data analytics prominent, professional sports began using statistics for decision making. Coaches and managers decide what players to play and when, what plays to run, make hiring decisions, and design game strategies all based on data. These data run along a chronological continuum from capturing moment-to-moment actions to examining long-term trends, just as in education. For example, a tennis player’s decision of whether to hit to the opponent’s forehand or backhand is based on data trends of stroke strengths and weaknesses but also on how the opponent is hitting at that very moment, a short-term set of data. A hiring decision for a new pitcher may be based on long-term data trends.

Bringing the metaphor full circle, colleges of education are also service providers. They serve their students by providing them with the knowledge, tools, and dispositions they need to be educators. They serve the schools and districts into which their teacher candidates and graduates are placed. It is critical for the educator preparation programs to ensure that their teacher candidates and graduates are sufficiently initially prepared to meet the needs of the districts by creating an essential and tight feedback loop that provides actionable data that inform a continuous improvement cycle in the EPP. These data may vary along a continuum of rapid feedback, providing information about excellence in or lack of preparation in particular areas (e.g., a content discipline, classroom management, etc.) of graduates that align with hiring needs.

Educator preparation has been the target of policy makers and advocacy groups for many years. The National Council on Teacher Quality (NCTQ, 2014) has been an outspoken critic about the lack of quality among American teachers and the preparation programs that feed into the U.S. schools. For better or worse, NCTQ has done ratings of educator preparation programs that have been met with disdain by colleges and universities and created significant pushback and dialogue in the field. Before former Secretary of Education Arne Duncan left office, he established regulations that would
attempt to improve teacher preparation by providing guidelines by which the programs should function (U.S. Department of Education, 2014).

As society changes, what people need to know and be able to do changes as well. The schools of today are not working for many students, even those who have been successful in the past. As schools change, EPPs also need to change the way that they are preparing teachers. EPPs must reflect societal changes and be able to adjust to emerging needs and trends. For example, as educational technology emerged in the 1980s, it was important that EPPs began to integrate technology into the curricula. A current trend is the growing importance of educators knowing how to use data effectively and responsibly (Data Quality Campaign, 2014; Mandinach & Gummer, 2013, 2016). Although EPPs recognize this need, many have not yet addressed this need (Mandinach, Friedman, & Gummer, 2015; Mandinach & Nunnaley, 2017). Other topics that have emerged that EPPs must address include personalized learning, formative assessment, school climate, bullying, student diversity, student mobility, and homelessness. EPPs must be malleable and responsive to enable them to develop educators who themselves can adapt to emerging needs.

The field of education and the general public recognize that improvements can and should be implemented, but there are legitimate challenges. Institutions of higher education are highly complex organizations with deeply embedded systems and structures that control their functioning. Even if there is the desire to make changes for improvement in preparation programs, the structures of the universities in which the programs are situated have the potential to present hurdles organizationally. Colleges of education are only one component in larger institutions, and desired changes in policies and practices in that one component may not resonate university-wide. For example, if a school of education wants to initiate practices that impact tenure and promotion among faculty, policy changes may need to be brought in front of the university’s entire faculty senate to align policies across colleges and departments. Further, the colleges of education must interact with other departments where their students may be receiving content-based courses, thereby necessitating the creation of feedback loops across departments. Many colleges of education receive students who have come from the community college system, where the students receive initial exposure to course content. Feedback loops also must exist between the programs and these feeder institutions to determine if the students have been adequately prepared for future coursework. Colleges of education serve to provide public and private schools with qualified teachers and administrators. There must be established relationships and another set of feedback loops to understand the impact of teacher preparation on practicing educators. Because colleges of education are closely responsive to other complex systems with components that are often beyond their control, sharing clear, explicit information about educator preparation in one system and its influence on practice in another system is essential. These are very real challenges.
Take, for example, a preparation program that finds its elementary teacher candidates are deficient in content knowledge, say in mathematics, once they are out in a school district as in-service teachers or as graduates. The school district may come back to the program and provide such critical feedback. If teacher candidates in the program receive their math content in the program itself, there may be some internal mechanism to provide a foundation for future improvement. However, if teacher candidates receive their content courses in an external department such as in mathematics, science, or history, then there is less immediate control for raising the problem and creating actionable steps toward improvement. The responsibility for improvement resides in multiple departments, and the extent to which a mathematics department may be receptive to that responsibility is a concern. As in K–12 education, where high school teachers blame middle school teachers who blame elementary teachers about what students know and are able to do, college faculty argue that prior preparation limits what they can achieve with students in their own programs. Further, is it actually the students’ poor knowledge of math, or is it their pedagogical content knowledge— that is, their knowledge of how the math can be transformed into instructional actions? How can the college of education work productively with the content departments to effect change?

Another challenge in effecting change among EPPs is the diversity of the institutions that prepare teachers. There are huge programs that produce hundreds of educators each year. These are typically the large, land-grant institutions. These institutions balance the need to excel in research and the production of educators. Faculty are highly specialized and are rewarded more for their research than for their teaching. There are the traditional state colleges that also produce many teacher candidates for local education agencies. The focus here may be more on developing educators who can serve the local districts. Then there are the many small, independent colleges in which a department of education is situated with only a handful of faculty members. Such departments may graduate a small number of students and have faculty who must teach a range of courses. Each type of institution presents different visions, opportunities, and challenges in the quest to use data for continuous improvement. In brief, there is no one-size-fits-all education program. There may be some foundational data that are common across all programs, but there may also be unique sources of data. Similarly, there may be common challenges across these programs but also challenges specific to the institution.

Other forms of programmatic diversity also must enter the discussion. There are virtual programs, with students and graduates dispersed across the country and even internationally. There are for-profit programs where emphases may be very different. And now an emerging trend is the proliferation of residency programs that have somewhat different models of delivery and focus, as well as different staffing models. In fact, some residency programs have developed incredibly strong institutional research foci to use
data as part of a feedback loop to improve their performance (Michael & Susan Dell Foundation, 2016a, 2016b, 2016c). It is possible that more traditional programs can learn from practices being implemented in the residency programs (Michael & Susan Dell Foundation, 2016d).

Despite the many challenges, there is no question of the need for data use for programmatic improvement. EPPs need actionable data on which to make decisions about how to improve. The data may come from the program itself, or it may come from a variety of sources, including feeder institutions, content departments, and school districts where teacher candidates and graduates are placed. Having the right data to answer the pressing questions is only one step in the improvement process. Other components are also needed to create embedded data cultures in the programs where data use is expected and the appropriate infrastructure and supports are provided. We now turn to the data literature in the K–12 sector from which parallels can be extended to data use in EPPs.

In parallel, the literature on improving schools at the K–12 level can also be informative to the focus of the book. Bryk, Gomez, Grunow, and LeMahieu (2017) outline the concept of networked improvement communities. Many of the principles generalize the EPPs. The systemic nature of the improvement process for EPPs can be considered a networked improvement community. This perspective recognizes the complexity of the system in which EPPs must function, that EPPs cannot and do not function in isolation, the centrality of measurement and appropriate metrics, the importance of data collection targeted at the change process, that there are key drivers that relate to the intended objectives, and that there are no easy solutions to complex issues. Improvement science will be discussed in more detail in Chapter 2.

Building a Data Culture in Educator Preparation: Parallels to Schools and Districts

Research in the area of educational data-driven decision making has long focused on the need to establish data cultures in which the use of data is deeply embedded in routine practice. Almost a decade ago, an extensive literature review was conducted by a panel of expert researchers and a practitioner to discern pervasive patterns and trends in the data literature from which recommendations could be made (Hamilton et al., 2009). The focus of this review was how data were being used at the classroom, school, and district levels to identify best practices. The review produced five recommendations for data practice in K–12 education to facilitate effective data use. Essentially, the recommendation laid out a roadmap for the establishment of data cultures. These recommendations can be examined to ascertain the extent to which they generalize to educator preparation programs, as they appear fundamental to good data use in education. The components underlie much of the content of the chapters in this book.
Ellen B. Mandinach and Edith S. Gummer

We transform the five recommendations into parallel principles in educator preparation.

1. “Make data part of an ongoing cycle of instructional and programmatic improvement.”
2. “Provide multiple opportunities for teacher candidates to examine their own data and set learning goals.”
3. “Establish a clear vision for college-wide data use.”
4. “Provide supports that foster a data-driven culture within the college or university.”
5. “Develop and maintain a cross-programmatic data system.”

(adapted from Hamilton et al., p. 9)

Fundamental to data use is the concept of an iterative cycle of inquiry or a process of feedback that leads to continuous improvement. In Senge’s (1990) organizational change work, feedback loops fed by data provide the necessary foundation to understand the mechanisms of the change process. EPPs share many commonalities with other organizations. They are learning systems, and staff should be supported to understand that they are part of an organization that must continually improve through self-examination. The typical inquiry cycle includes: (a) the identification of a problem of practice; (b) the collection of relevant and actionable data; (c) the analysis of the data that leads to an interpretation and hypotheses about the root causes of the issue; and (d) the modification of practice, based on the hypothesis that a particular intervention will address the problem of practice. But the continuous improvement efforts are not complete at that point. It is not a simple linear process with a solution at the end. It is iterative. Once Step (d) has occurred, more data will be needed to determine if the intervention has had the desired impact or if other modifications need to be taken. Educator preparation programs always need to be testing the extent to what they are doing are aligned with the evolving systems in schools and districts, so the cyclical process of improvement truly is one of continuous improvement.

The second component promotes the notion that teacher candidates should become their own data-driven decision makers. They should be part of the data collection process. The recommendation includes specific steps such as communicating clear expectations and assessment criteria, providing concrete feedback, providing tools to learn from the feedback, and using the data to guide next action steps. These concepts are readily transferable to educator preparation. The use of their own data can be modeled for teacher candidates so that they develop practices they will use to help their own students do the same once they are in the classroom. Teacher candidate data often are at the heart of the metrics and indicators of success that are used for program evaluation, including observations of teaching and the production of artifacts through assignments and assessments. So helping teacher candidates to self-assess can become a critical part of the
improvement process. Chapter 7 discusses how the Relay Graduate School of Education uses this very premise as a key concept in their improvement process.

The third component of effective data use is the establishment of a vision for why data are to be used. This vision comes from leadership and must be explicit. Whether a principal, a superintendent, or a dean, having a clear message about the importance of and expectations for data use is essential. As Earl and Katz (2006) note about K–12 practice, data use begins and ends in the principal’s office. This concept is readily generalizable to higher education. A dean who is a proponent of data use will be much more likely to inspire and require that faculty and staff be so inclined. Creating the vision includes some concrete steps. One is that the dean should lead with data. Leadership is essential, but the dean will need help. Therefore, a distributed model of leadership is recommended, one where there is a shared level of responsibilities. A second component is to establish a data team within the organization that models and promotes the use of data. These would be lead faculty and staff that are part of the educator preparation program. A third premise is that critical teaching and learning concepts must be defined. For example, the InTASC standards create a skills load comprised of performance, essential knowledge, and critical dispositions across 10 standards that is unlikely to be measureable (CCSSO, 2013). The issue is not having teaching and learning concepts but rather about prioritizing them and identifying the developmental constraints of initial preparation. Metrics and indicators of success are fundamental concepts that resonate across the chapters in this book. People need to understand how they are being measured. Finally, there needs to be an articulation of roles and responsibilities. If faculty and staff are being expected to collect, examine, and act on data, they need to understand and agree to what they are being asked to do with sufficient support and authority to make changes. Managing the change environment needs to be a crucial aspect of moving to more effective use of data to inform, improve, and innovate teacher education programs.

Resources are needed to deeply enculturate the use of data in an organization. Although data use must become a routinized practice, there is no question that, for some, it will be an additional, new activity, and even an added burden, one with which some may feel uncomfortable or unfamiliar. Having a data leader or a data coach can ease some of the tensions. These are individuals who know how to use data within the context of the organization. They are not just statisticians or data nerds. The more distributed this leadership, the better. A second resource is for there to be dedicated and structured time for faculty and staff to collaborate around data use. As we see in Chapter 9, Texas Tech uses what they call Data Days. Peck and Davis in Chapter 4 refer to Data Summits. Educator preparation programs have begun to dedicate time and days to the examination of data. Such meetings provide time for the collaborations and in-depth conversations that are needed to examine, interpret, and act on the data. A third
component is ongoing professional development. This is certainly necessary in the K–12 setting. It is not clear that faculty and staff in higher education are really well versed in data use; that is, they may not be the most data-literate individuals. Some people may be the typical data gurus. Most, however, could benefit from learning how to use data in an effective manner and co-constructing ways to do so in their own programs. They also may need to receive instruction on how to protect the privacy and confidentiality of student data, a growing concern in both the K–12 and higher education arenas. A final consideration is the extent to which there is a need to provide incentives to faculty and staff for their data work. If additional work is involved, leadership needs to consider the impact there might be to contracts, tenure, promotion, and workload.

The final recommendation needed for an effective data culture is the provision for a data system that supports the kinds of data to be collected and the vision for data use. The metrics and indicators will help determine how the data are collected and accessed. In educator preparation programs, given the nature of the collaborations necessary to obtain the relevant data, this may mean that the systems have to collect data from disparate sources, including participating districts and collaborating departments within the university. Memoranda of understanding (MOUs) are therefore essential. A number of chapters in this book address the need for appropriate data systems, particularly Chapter 8, with a focus on Arizona State University.

Overview of the Book

This volume addresses the complex issue of how educator preparation programs are or are not using data to inform their efforts to continually improve. It explores the policies and practices that influence colleges and departments of education. It examines the increasing pressures and emphases to improve institutional functioning to produce quality graduates ready to enter the teaching profession, recognizing that programs function amid complex systems with structures and organizations such as accreditation agencies and standard-setting bodies on one end of a continuum of drives for improvement and schools and districts as the recipients of their graduates at the other end. The authors of the chapters employ different methods to support data use and are addressing different aspects and problems of their teacher preparation programs. They have focused on their lessons learned and provide good recommendations for readers.

The book is separated into four sections: an introduction; chapters from researchers; chapters from the field of practice; and a concluding chapter.

In Chapter 2, Gummer and Mandinach provide a landscape view of the continuous improvement process, especially as it pertains to educator preparation programs. This chapter lays the groundwork for why an orientation for continuous improvement is needed in colleges of education and educator preparation programs, more specifically. It discusses the policies
and the pressures for programs to improve and to become more responsive to the needs of districts. The chapter examines trends in policies that regulate preparation programs. It addresses how data use can be employed to mitigate the critics of the teaching profession. The chapter reviews existing research and highlight some best practices. It draws parallels to other professionals where cultures of improvement and data use already exist and extract lessons that may be applied in education.

In the first of the research-oriented chapters (Chapter 3), Michael Allen and Charles Coble, from Teacher Preparation Analytics, review state policies and roles in the educator preparation improvement process. They discuss the impact of national professional organizations. Given their vast experience working with states and organizations, they draw on their experience to provide salient examples. Allen and Coble provide a landscape view of the programmatic indicators and discuss how these metrics can be used. They provide a realistic discussion of the challenges, risks, and limitations to this work. Chapter 3 concludes with a view to the future and steps forward to make the continuous improvement process possible.

In Chapter 4, Charles A. Peck of the University of Washington and Susannah Davis of Oregon State University draw from research and policy on data use in teacher education. They provide a holistic view of how tools, people, and organizations impact the continuous improvement and organizational change process. They provide targeted recommendations on how institutions can build the capacity to use data for continuous improvement. In this chapter, Peck and Davis review the current knowledge base related to the implementation of data use policies. The authors focus particularly on the kinds of organizational policies and practices that have been developed within teacher education programs that have “leaned into” the challenges of data use (Davis & Peck, 2017; Peck & McDonald, 2013). By this, the authors refer to programs that have proactively and strategically responded to emerging data use mandates in ways that reflect their commitment to local program values and integrity without abandoning their responsibilities for public accountability. Peck and Davis’s analysis of the emerging research literature on these programs suggests that effective implementation of data use policies, particularly the development of meaningful and productive data use practices, requires orchestration of organizational learning resources that operate at multiple levels, including those related to the choice of data sources, personnel policies, communication practices, and, above all, leadership vision and skill. The chapter concludes with directions for practitioners and policy makers aimed at making data more useful and more widely used in the improvement of teacher education programs.

The Deans for Impact is a nonprofit organization that brings together member educator preparation institutions to seek solutions to transform the way educators are prepared. In Chapter 5, Tracey Weinstein and Charis Anderson present the organization’s four guiding principles that drive the design of educator preparation programs. The authors outline the indicator
system and strategies used by member institutions for improvement. They draw on specific examples to highlight the strategies and guiding principles. This chapter explores the collection and use of evidence of candidate progress and performance across an institutionally diverse group of educator preparation programs—private, research-intensive universities; public land-grant colleges preparing large numbers of educators; urban teacher residences; and new, less-traditional programs. The chapter highlights common challenges and opportunities identified through this work and proposes a set of steps intended to enhance the effective use of data by educator preparation programs.

In Chapter 6, Ana Quintana, Bryan Richardson (formerly) of UPD Consulting, and Cari Reddick share research that has led to the development of technology to create the Teacher Preparation Data Model (TPDM). The TPDM is an extension to the Ed-Fi Data Standard for teacher preparation use cases that will enable the appropriate exchange of data between organizations in the teacher preparation domain for continuous improvement and connect that data to K–12 systems. The Ed-Fi Data Standard is the widely adopted, open-source data standard developed by the educational community for the betterment of the community. The Ed-Fi Data Standard serves as the foundation for enabling interoperability among secure data systems and contains a unifying data model designed to capture the meaning and inherent structure in the most important information in the K–12 education enterprise. The objective of the TPDM is to facilitate the integration of data and data use in teacher preparation through a cost-effective and scalable model. The chapter describes the development of TPDM and how the TPDM introduces an opportunity for teacher preparation programs, local education agencies, and state education agencies to explore how the data model might improve the preparation of their teacher candidates by becoming adopters of the data standard and its associated data integration tools.

In Chapter 7, Kaycee A. Salmacia and Liam Honigsberg of the Relay Graduate School of Education present the story of how the institution has used data to transform its practice. Relay uses an indicator, Student Growth and Achievement (SGA), as a metric for success of its students and its faculty. This chapter describes the success and obstacles Relay has encountered as it has introduced the SGA and trained personnel to use the metric. The chapter recounts that many of Relay’s inaugural faculty had not measured student outcomes in their own tenure as a K–12 teacher and thus lacked comfort with the idea of teaching others how to do something they had never done. Relay’s first technological tools for storing and exploring data were often brittle and difficult to use, which caused frustration among graduate students and took time away from other important teaching duties. The Relay chapter is the story of how the institution applied the principles of outcome-driven data literacy to solve its problems and to turn its content from being a major challenge and problem for faculty and students to the content nearly 100% of faculty have embraced.
In Chapter 8, Edith Gummer describes how the Mary Lou Fulton Teachers College (MLFTC) at Arizona State University is transforming itself through the use of data. The MLFTC has been evolving a sophisticated data collection and visualization process since 2010. Grounded in the university’s student information system (PeopleSoft) and the candidate placement and assessment system (Watermark, formerly Tk20), the iTeachAZ warehouse, dashboard, and mobile application collects and provides appropriate displays of data for multiple stakeholders. These include administrators, site coordinators/university supervisors, mentor teachers, and teacher candidates. The iTeachAZ mobile application facilitates collection of data around student performances when supervisors do not have access to the Internet and provides the collection and submission of data that different stakeholders have around concerns for candidates that need to be addressed by advisors and other support mechanisms. Much of the use of the data is for monitoring and compliance purposes to ensure that adequate progress is being made in support and supervision of candidates. At the level of site coordinator/university supervisor interactions with teacher candidates, the data and displays structure and authenticate the feedback that candidates get of their performances in their field placements. However, little of the multiple types of data are used specifically for continuous improvement at the program level. As MLFTC engages in improvement and innovations, specific processes are being designed to ensure that program coordinators and faculty have ways in which to focus their decision making around data from their courses and students.

Chapter 9, the final chapter from the field, provides lessons learned from the work that D. Scott Ridley and his colleagues at Texas Tech University have experienced. Ridley and Hamman present five lessons learned through their efforts at transformation and how deeply embedded this work is in relation to cooperating districts. The lessons are both based on extensive experience working with the districts and internal staff and forward thinking about how to make the improvement process work in the future. Texas Tech’s lessons are highly generalizable and should resonate across educator preparation programs. Although these lessons are articulated in hindsight, they may provide a “bootstrap” for those who are ready to respond and confirmation of those currently engaged in the struggle.

In Chapter 10, Mandinach and Gummer review the findings and take away messages culled from the contributing authors. They discuss the challenges and opportunities (the CHOPS) that are drawn from the chapters. Challenges always balance the opportunities. In many instances, one overarching opportunity outweighs the many challenges. That opportunity provides the motivation for change. It stimulates institutional growth and learning. In the case of data use in educator preparation programs, the opportunity is for programs to use data to inform their practice and to stimulate continuous improvement in the delivery of curricula and practical experiences aligned to the emerging needs of the U.S. educational system.
The chapter is grounded in the notions of systems thinking, noting the interplay among the complex and sometimes competing organizations internal and external to educator preparation programs. The chapter then synthesizes lessons drawn from the previous chapters. It discusses the many challenges that these cutting-edge programs are facing and extracts processes that are being shown to be effective. The chapter attempts to generalize how these processes may be implemented in other institutions, noting their diversity and the complexity of such an enterprise. The chapter concludes with an action plan, a delineation of the actionable steps based on the use of data that educator preparation programs can implement to create data cultures that support evidence-based continuous improvement.

References


What Is Continuous Improvement?

The term “continuous improvement” is used across industries to describe a process or approach to problem solving that represents an ongoing effort to improve outcomes (American Society for Quality, n.d.). In education, Park, Hironaka, Carver, and Nordstrum (2013) characterize continuous improvement as “the act of integrating quality improvement into the daily work of individuals in the system” (Park et al., 2013, p. 5). They differentiate continuous improvement from improvement science, the process of ascertaining what works, for whom, and under what conditions, and from quality improvement, which they characterize as “the disciplined use of evidence-based quantitative and qualitative methods to improve the effectiveness, efficiency, equity, timeliness or safety of service delivery processes and systems” (Park et al., 2013, p. 7). In continuously improving systems they argue that change can occur both quickly and incrementally, as organizations learn from experience while testing and refining strategies to produce better results. In education, continuous improvement can refer to a school, district, or other organization’s ongoing commitment to quality improvement efforts that are evidence based, integrated into the daily work of individuals, contextualized within a system, and iterative (Park et al., 2013).

Continuous improvement involves a cyclical approach to problem solving: it allows relevant actors to reflect on their work, identify problem areas, pilot potential solutions to those problems, observe and evaluate interventions, and adapt interventions based on data collected (Flumerfelt & Green, 2013). Many models of continuous improvement include very similar characteristics and components, but it is important that practitioners and designers know the similarities and differences, particularly in terms of the types of problems addressed, how those problems are identified, how solutions surface and are tested, and how findings about the improvement are dispersed (Bryk, Gomez, Grunow, & LeMahieu, 2015; LeMahieu, Bryk, Grunow, & Gomez, 2017).

Continuous improvement has been used in multiple industries, businesses, and professional practices. The Institute for Healthcare Improvement (IHI,
The Continuous Improvement Process

2003) has conducted over 50 collaborative projects in short-term projects to improve a focus area to reduce problematic practices in hospital waiting times, decreasing ICU costs, and shortening hospitalizations for patients with congestive heart failure. Other IHI initiatives, such as the Pursuing Perfection Initiative (Kabcenell, Nolan, Martin, & Gill, 2010), have addressed innovative solutions to reducing mortality and better patient experiences, though decreasing health care costs remained elusive. Continuous improvement efforts in the legal profession have documented improvements in the flow of the review of documents, improving indexed compilations of documents for future reference, and increasing the efficiencies in the ways they interacted with clients (Clifford, 2014).

Why an Orientation for Continuous Improvement Is Needed in Educator Preparation Programs (EPPs)

There are multiple reasons why an orientation for continuous improvement is needed in EPPs. From a business perspective, universities compete for students, especially in current times when enrollment in EPPs is declining. The US News and World Reports ranks graduate and online graduate education programs, and the competition to improve in rankings takes valuable time and energy.

The teachers who graduate from EPPs express varying levels of satisfaction with the quality of their EPP (Darling-Hammond, Chung, & Frelow, 2002). These areas of dissatisfaction influence the extent to which teachers felt prepared to teach in the classroom (Beck, Kosnik, & Rowse, 2007; Bruneau, Hoz, & Silberstein, 1997; Schulz, 2005). In a study on novice elementary school teachers, beginning teachers expressed discontent in continuing to teach because “there was a disparity between their teacher education programs and the ‘real’ teaching world” (Barrett Kutcy & Schulz, 2006, p. 78). The authors found teacher preparation programs did not prepare new teachers for such disparities nor were they equipped to teach in their first classrooms. Curriculum developers of teacher preparation programs emphasize the theoretical component, and theoretical courses may lack the key components that can improve a teacher’s practice and student learning (Beck et al., 2007; Brzycki & Dudt, 2005; Schulz, 2005). Melnick and Meister (2008) reported “Doing school” cannot be simulated in the university classroom, and one intensive field experience cannot equip preservice teachers with the essentials to succeed in their own classroom” (p. 53).

Najal (2010) indicated that 100% of the participants in the qualitative study of the perspectives of first year teachers indicated that EPP theoretical courses were irrelevant for classroom teaching in the first year and did not tie theory into practice. Only 75% of the beginning teachers indicated that the pedagogy skills learned in methods classes were useful for them in their classroom. All of the beginning teachers indicated that the classroom management knowledge and skills they gained from EPP classrooms were
not sufficient. A large majority (90%) indicated that they wished they had been given more practical activities relevant to classroom teaching such as ready-to-use strategies and real-world applications.

Policies and Pressures for EPPs to Improve and Become More Responsive to the Needs of Districts

The federal Department of Education had an initially strong voice in supporting accountability for EPP performance, with implications for continuous improvement. The department’s (U.S. Department of Education, 2016) teacher preparation regulations called for transparency around the effectiveness of all preparation programs (traditional, alternative routes, and distance). They indicated that states would be required to report annually, at the program level, on measures that included placement and retention rates of EPP graduates over their first three years of teaching, feedback from EPP graduates and their employers that rated the effectiveness of the EPP, and results from graduate evaluations in their teaching placements that included the achievement growth of their students and the results of teacher evaluations or other state-determined measures that meaningfully differentiate amongst teachers. Other, less specified, measures that provide evidence of the rigorous outcomes were also allowed. However, in the summer of 2017, Congress rescinded the Education Department’s rule under the Congressional Review Act, which allows Congress to disapprove regulations enacted toward the end of a president’s administration.

EPPs in the United States are operating in a context of increasing accountability policies at both the state and federal level. Continuous improvement is central to the new and intensified standards for evidence-based accountability and program improvement now integral to national standards for teacher education program accreditation and state program review and approval processes. These state and federal policies within the field of teacher education increasingly require outcomes-based evaluation, external accountability and oversight, and evidence of data-informed program improvement (Cochran-Smith, 2005a, 2005b; Cochran-Smith & Villegas, 2015; Wineburg, 2006). The seminal paper, Measuring What Matters: A Stronger Accountability Model for Teacher Education (Crowe, 2010), articulates the cogent argument for states to build better accountability systems to drive toward improved quality in EPPs. Crowe laid out the dimensions of a state accountability system that focused on more common indicators, determining where and how long teachers work in the profession and, ultimately, the impact of teachers on student achievement. The states were recommended to focus on higher standards for admittance, better monitoring of field experiences, and stronger program evaluation based on important outcomes.

Efforts by the Council for State School Officers (CCSSO) in their report written for the Task Force on Educator Preparation and Entry into the Profession (CCSSO, 2012) laid out the key drivers in state policy that might
be brought to bear to improve EPPs. The elements of a “learner-ready” teacher included characteristics of modeling for and developing in students the ability to think critically and creatively, to be able to solve real-world problems through deep knowledge of content and pedagogy, to develop literacy across subject areas, to collaborate and work in teams, and to guide students to take responsibility for their own continuous learning. Teachers accomplish this by understanding the differing needs of students and personalizing learning, by holding them to high expectations, by deeply engaging students in learning, by using data to ascertain student progress and adapt instruction, and by focusing collectively on continuous improvement. These characteristics of “learner-ready” teachers are accompanied by those for “school-ready” principals. The report contains recommendations to states for licensure; program approval; data collection, analysis, and reporting.

Additional work by the CCSSO in their Network for Transforming Educator Preparation (NTEP) has resulted in multiple reports that further articulate what it is that EPPs need to do to improve their experiences for preparing teachers. Together with the Collaboration for Effective Educator Development, Accountability, and Reform (CEEDAR), they have continued to provide information for the states to guide them in developing systemic responsibility for reforming education (CCSSO, 2018a). A second action group from the CCSSO task force focused on the design of the outcomes-oriented data systems for EPPs (CCSSO, 2018b). These systems included the capacity to ensure that data could be reported back to EPPs so that they could use the information for continuous improvement, provide a mechanism so that the EPPs could be transparent and publicly report their progress in accountability, and provide the states with information for approval and renewal decisions. CCSSO revisited the focus on measuring what matters with a second document in 2018, describing the processes that the six participating states have implemented that provide strengths and limitations of current data collection mechanisms (CCSSO, 2018b). The document clearly identifies the need for continued development of data collection processes in order to provide better evidence that might be used to inform continuous improvement efforts.

State licensure and EPP certification regulations are one lever by which continuous improvement efforts are driven. Additional influence is found in the work of national accreditation processes. The Council for the Accreditation of Educator Preparation (CAEP) clearly states a focus on evidence-based decision making and continuous improvement at the organizational level in one of its two primary guiding principles as they introduce their standards with the statement: “There must be solid evidence that the provider’s educator staff have the capacity to create a culture of evidence and use it to maintain and enhance the quality of the professional programs they offer” (CAEP, n.d.; http://caepnet.org/standards/introduction).

The CAEP standards to which teacher education programs are held include more detailed requirements related to evaluation of programs’ impact
on P-12 student learning and continuous improvement at the level of the EPP in Standard 5, indicated next.

The provider maintains a quality assurance system comprised of valid data from multiple measures, including evidence of candidates’ and completers’ positive impact on P-12 student learning and development. The provider supports continuous improvement that is sustained and evidence-based, and that evaluates the effectiveness of its completers. The provider uses the results of inquiry and data collection to establish priorities, enhance program elements and capacity, and test innovations to improve completers’ impact on P-12 student learning and development.

(CAEP, 2016, p. 2)

The specifics of continuous improvement for CAEP Standard 5 continues with the following:

5.3. The provider regularly and systematically assesses performance against its goals and relevant standards, tracks results over time, tests innovations and the effects of selection criteria on subsequent progress and completion, and uses results to improve program elements and processes.
5.4. Measures of completer impact, including available outcome data on P-12 student growth, are summarized, externally benchmarked, analyzed, shared widely, and acted upon in decision-making related to programs, resource allocation, and future direction.
5.5. The provider assures that appropriate stakeholders, including alumni, employers, practitioners, school and community partners, and others defined by the provider, are involved in program evaluation, improvement, and identification of models of excellence.

(CAEP, 2016, p. 2)

CAEP with support from Pearson Teacher Education commissioned Teacher Preparation Analytics (TPA) to articulate the “state of the art” processes determine the quality of EPPs across the country in order to better understand the characteristics of EPPs around which data should be gathered. (Allen, Coble, & Crowe, 2014). Presenting the Key Effectiveness Indicators (KEI), the authors provide the framework for effectiveness in EPPs from the perspective of teacher candidate characteristics, the knowledge and skills for teaching they should demonstrate, the performance they should show in the classroom, and what employment and retention in teaching positions should address, with a special emphasis on high-need areas. TPA (2016) consequently produced a guide to the KEI which builds on the report by the National Academy of Education report on methodology for EPP program evaluation (Feuer, Floden, Chudowsky, & Ahn, 2013). The guide lays out
the indicators and measures that are appropriate for each category of assessment embodied by the KEI and identifies strengths and weaknesses.

**Existing Examples of Continuous Improvement in EPPs**

The current descriptive research literature on examples of continuous improvement practices by EPPs is relatively thin, especially when compared to the efforts that have been carried out in K–12 education. The authors of the chapters in this book are evidence of growing practice. Several examples from the research literature, not including those carried out by other authors of chapters in this publication, are offered here to indicate the sorts of practices and contexts in which continuous improvement in EPPs is being conducted, but the examples are not intended to be encyclopedic.

Colby and her colleagues (Colby, Lambert, & McGee, 2016) examined the use of Collaborative Analysis of Student Learning (CASL) as a mechanism to engage in continuous improvement during an accreditation cycle. The focus of their study was to determine the ways in which faculty were using collaborative analysis to structure their analysis of student learning and identify ways to improve their programs. They emphasized the importance of administrator and faculty commitment to the process and identified the need for materials to structure the review of student work. Findings indicated that faculty discussed areas of course development, curricular changes, and student products. Action items included improving alignment of documents and processes to standards, revision of assignment guidelines and rubrics, and revisiting standardized rubrics for student work. Student product action items focus on ways to improve deeper critical thinking of their students, identifying ways to integrate skills into student products and ways of helping students with deficiencies.

Snow and her colleagues at Boise State University (Snow, Dismuke, Zenkert, & Loffer, 2017) describe a shared leadership model wherein a small team engaged with an associate dean of education to address quality assurance concerns raised by a prior EPP program review and accreditation study. The initiatives piloted included the use of a new data system (TaskstreamTM), the requirement of a standard teacher performance assessment used by all programs (S-PAT), and new final clinical experience assessments. These data were used to focus the faculty on the CAEP standards as well as to build a community of inquiry around the outcomes of the EPP programs. Using a focus on professional capital (Hargreaves & Fullan, 2012) and social networks in education (Daly & Finnigan, 2010), the authors identified how evidence-informed decisions were negotiated between a central team and the rest of the faculty. This case study of the efforts of the elementary education program continuous improvement process demonstrates what the authors refer to as “decisional capital,” including the ways that participants could influence decisions and gain empowerment as they took collective responsibility for decisions (Snow et al., 2017). Focusing the work of
the team on collaborative analysis of data provided for the development of professional capital by the participants, with a more nuanced and deeper understanding of the role of data rather than impressions to inform decisions. Balancing the focus on teacher candidate performance data and the underlying principles shared by the faculty became an important element of the work of the data team. Sharing responsibility for making decisions across a larger group emerged as a beneficial element for the program. The authors indicated the importance of the CAEP standards to guide their evidence-informed decisions. Having a data system enabled the team to better manage and understand the implications of the data they were collecting and using for decisions.

Continuous Improvement Models

The Carnegie Foundation for the Advancement of Teaching has been focused on the improvement of education, with a special focus on improvement science under the leadership of Anthony Bryk (Bryk, Gomez, Grunow & LeMahieu, 2015). In a recent issue of Quality Assurance in Education, these and other authors lay out the similarities and differences of seven different approaches to improvement science in education including networked improvement communities (NIC), design-based implementation research, Deliverology, implementation science, Lean for education, Six Sigma in education, and positive deviance (LeMahieu et al., 2017).

A networked improvement community (NIC) connects a focus on education systems as networks of individuals within and across organizations who are engaged in identifying important persistent problems of practice and iteratively identifying and testing out solutions (LeMahieu, Grunow, Baker, Nordstrum, & Gomez, 2017). Carnegie’s networked community focus is based on the work of Engelbart (1992) that organizes individuals as practitioners, organizations internally, and cross-organizational systems. The four stages of the Carnegie focus on improvement—Plan, Do, Study, Act (PDSA)—illustrate continuous improvement process based on the work of Deming (1994). In the Plan stage, a continuous improvement team studies a problem that needs to be solved, collects baseline data on that problem, elaborates potential solutions to that problem, and develops an action plan. In the Do stage, the team implements its action plan, collects data on its intervention, and records developments. In the Study stage, the team gauges the success of the intervention by comparing baseline and new data, analyzes results, and documents lessons learned. In the Act stage, the team determines what to do with its results. Depending on the success of its intervention, the team may choose to adopt, adapt, or abandon its tested solution (Langley, Moen, Nolan, Nolan, Normal, & Provost, 2009). With PDSA, as in other continuous improvement models, the use of an iterative process to rapidly move potential solutions into practice allows for ongoing quality improvement. After completing one cycle, the team begins a new cycle to test different solutions or address new problems.
In design-based implementation research (DBIR), a relatively recent model of improvement science, the focus is on persisting problems of practice, a collaboratively engaged iterative focus on designing and testing solutions, a commitment to systemic inquiry that emphasizes both the development of theory and knowledge about processes for program implementation as well as measurement of learning outcomes, and a focus on organizational capacity development to sustain improvements in educational systems (LeMahieu, Nordstrum, & Potvin, 2017). Penuel and his colleagues emphasize the focus on the classroom practitioner as the driver for the implementation design research (Penuel, Fishman, Cheng, & Sabelli, 2011).

Deliverology shifts the focus from the problems of practice at the practitioner level espoused by Penuel and colleagues (2011) to a focus at the level of education organizational leaders and their goals and planning processes (Nordstrum, LeMahieu, & Dodd, 2017). The focus at the organizational level is on what the system is trying to do (goals), how they will plan and implement the changes, what evidence they will collect about the progress they are making, and the correction mechanisms they will take if they are not on track for success. The emphasis is on a smaller unit within an organization being responsible for working with leadership to achieve results.

Implementation science has a longer history in other disciplines that focuses on the extent to which interventions and programs are implemented with fidelity, recognizing that behaviors, beliefs, and values of practitioners have significant influence on eventual success (Kelly, 2012). In education, the focus of implementation science is on ensuring that the individuals and groups that are the target of improvements, typically in the form of interventions, are adequately supported during the improvement process. Designers, researchers, and policy implementers are equally responsible for helping insure that an intervention is successful (Nordstrum, LeMahieu, & Berrena, 2017).

The Lean improvement focus came most directly from industry and has gradually transitioned into education. The historical focus on ensuring that participants in improvement efforts are considered and engaged has the potential to make a Lean focus appropriate for educational systems improvement consideration. Similar to the process of Plan-Do-Study-Act in the Carnegie NIC process, the Lean methodology is cyclical in nature, with a focus on Plan-Do-Check-Act. As with Deliverology, the organization of the Lean methodology appears to be with a small design team that focuses on process improvement under the auspices of a leader (LeMahieu, Nordstrum, & Greco, 2017).

Similar to Lean, the origin of Six Sigma is in business and industry. The focus on improvement is on enhancing organizational processes of doing work by decreasing variability and eliminating activities that do not contribute to improvement efforts. The name “Six Sigma” relates to the “defects per million opportunities” that became the target performance for variation in products (LeMahieu, Nordstrum, & Cudney, 2017, p. 92). The methodological focus is on Define-Measure-Analyze-Improve-Control (DMAIC),
and practitioners collaborate with staff and administrators to identify specific problems and identify interventions for improvements.

The final improvement science model discussed in the Quality Assurance in Education issue is that of Positive Deviance. This process approaches improvement of problems from a different perspective than those indicated previously by operating on the assumption that the solutions to particular problems of practice already exist within the system and are being practiced by some individuals. These people are identified as “positive deviants,” and it is their identification and the study of what they are doing that is the primary work of the improvement efforts. The steps of the Positive Deviance process include defining the problem and intended outcomes, determining common practices and uncommon successful strategies, designing an intervention based on those uncommon successful strategies, monitoring progress, and disseminating results (LeMahieu, Nordstrum, & Gale, 2017). This process seems similar to the efforts of a number of organizations and philanthropies to help educators see examples of positive school systems and cultures as a part of their work.

Common to all of these models of improvement science is a focus on evidence that is iteratively interwoven in the improvement efforts. Forms of data that are brought to bear in these models are wide ranging, from quantitative statistical analysis of budgets and measurable outcomes such as attendance and student achievement to qualitative descriptions of strategies and stories for success. While the elements of these models are similar, the differences in the theoretical focus and processes indicate how varied continuous improvement efforts can be. Many of the examples in the Quality Assurance in Education issue come from the K–12 world, and the potential for implementation in the environment of higher education or other educator preparation programs is still to be demonstrated.

**Incremental Improvement versus Reform and Innovation**

The final section of this chapter addresses the distinction between continuous improvement, which is frequently seen as incremental or evolutionary, and EPP reform or innovation, which is frequently seen as revolutionary (Duncan, 2009). We raise this issue because the practices that underscore continuous improvement all start from the given conditions and move them toward some explicated improvement. By its very nature, innovation in EPP suggests that we are working to envision a future we haven’t yet encountered, which means that the goals to which we aspire are fuzzy and the nature of the data we need to collect are still somewhat to be determined.

Duncan (2009), in his address at Columbia Teachers College, raised a theme that he espoused multiple times in his tenure as the U.S. secretary of education—that EPPs in America were doing a “mediocre” job of producing educators. He laid the groundwork for the need for revolutionary
change on the need to move from past and current education systems that are barely responsive to current needs to prepare children for work and life. He described the changing demographics of the country, which makes equity in education opportunities ever more challenging. Finally, he pointed to the aging of the teaching workforce whose replacement is not being met by current recruitment and retention of new teachers. He argued that one consequence of these three influencing conditions was the need to increase the expectations of how we educate educators.

because we ask much more of teachers today then even a decade ago. Today, teachers are asked to achieve significant academic growth for all students at the same time that they instruct students with ever-more diverse needs. Teaching has never been more difficult, never more important, and the desperate need for more student success never been so urgent.

(Duncan, 2009)

Duncan went on to quote Levine’s (2006) study of teacher education, to identify many of the concerns that are still raised in journalistic circles today. He invoked both the focus of the Race to the Top criteria and to Louisiana’s model of EPP accountability tied to the performance of graduates of EPP programs. While both of these examples have been roundly criticized, the focus on data within them is evident.

At about the same time, Levine (2010) published an article in Phi Delta Kappan in which he raised many of the same issues about what drives the need for change in education—changes in the economy, demographics, globalization, technology, and growing instances of children with disabilities. He spoke of revolutionary changes in society that would need to be paralleled by changes in EPP. He pointed out multiple lessons learned through the work of the Woodrow Wilson Teaching Fellowship and identified a number of recommendations that he felt would support revolutionary change in EPPs.

We would argue that we are experiencing two revolutions in EPPs. Duty and Kern (2014) articulate a vision and a process for those who are charged with bringing about change in education. They lay out the argument that radical change must happen within the culture of organizations and that the culture must address improving the systems that currently exist, at the same time visioning, designing, and prototyping to establish the conditions for the systems of education that we need for the future, as is shown in Figure 2.1. The diagram indicates that really thinking of revolutionary change requires that we recognize that we are operating outside of the solid set of recommendations that the committees and panels described in this chapter set in front of us. Those recommendations describe the foundations on which innovative EPPs should be established. But the data that we need to inform visioning may not yet have emerged from past work, and the process to
develop evidence is speculative at best. Cochran-Smith (2009) emphasized this when she wrote about the “re-culturing” of education and her work on reclaiming teacher education accountability raises many of the same culture concerns (Cochran-Smith et al., 2018).

Most of the examples of continuous improvement practice that are described in later chapters in this book are focused in improving aspects of programs that are relatively mature or are operating within a relatively normal range of educator preparation practices. Whether the clinical experience model is a single semester or a yearlong residency, the focus is largely on developing educators to operate in a 1 teacher : 1 classroom context. Thus, the programs are in the realm of “improving the system we have” category focused on what we are doing now. Traditional (and some alternative) EPPs focus on the completion of a relatively set suite of courses that results in a degree and recommendation for certification. Alternative EPPs are frequently partnered with a college or university so that educator candidates can receive state licensing and frequently an advanced degree as a result of participating in that program. Improvement activities rarely are directed to the “what’s next” category of Figure 2.1. Much of that work was carried out in the design phase before the alternative program was launched or under consideration during early implementation phases.

What do improvement efforts look like if we are seeking to be truly innovative in educator preparation? What are the demands on the nature of evidence when we are trying to inform design characteristics of something that is really future oriented? Consider an ambitious effort to not just improve multiple EPPs within a university but to innovate them that is the vision of educator preparation that is currently helping to structure

---

**Figure 2.1 Two revolutions for change in education**

(Permission for use requested from 2Revolutions.)
the redesign efforts at the Mary Lou Fulton Teachers College (MLFTC) at Arizona State University. The reform agenda is ambitious, designing a suite of “un-programs” that prepare educators to work in teams in schools and that move teachers and schools away from ratio of 1 teacher : 1 classroom of 30 students. The design is working to move away from the sort of EPP planning that is influenced by thinking of students in a particular program such as elementary education or special education to consider what competencies (knowledge, skills, and dispositions) that might be represented in a classroom more broadly. Key questions involve how we think differently about the multitude of standards and indicators that structure the work that colleges of education currently support, what the financial models of workforce for schools and districts indicate of what is feasible now versus how they might need to change in the future, what the roles of educators might be that would work in a teaming environment more similar to nursing than to current EPPs. And ultimately, we need to reinvent the nature of the data or evidence that support design decisions such innovation in EPPs will require.

In trying to personalize educator preparation, to change the way the college and its school and district partners think about the educator workforce, Dean Basile is organizing a systems perspective with initial efforts being spearheaded by MLFTC’s Division I. As shown in Figure 2.2, the areas of innovation range from reconsideration of the competencies that are necessary for beginning teachers versus those that are indicative of experienced educators, what individualized pathways to demonstrate those competencies might look like, what reform of the educator preparation curricula are necessary, what the nature of the field experiences would entail, the expansion of what specializations EPP candidates might want, and an expanded identification of the communities of educators with whom the university might partner. This innovation effort focuses on what might be, what we need to address to have a better educator workforce in the future.

While ASU has a strong data system in the use of Tk20 and the iTeachAZ data warehouse, dashboard, and mobile app (as described in Chapter 8 in this volume), the nature of the data currently being collected will not be sufficient for this grander vision. Continuous improvement, EPP reform, and EPP innovation are not frequently talked about as a continuum, and identifying how to examine EPPs as a system is not well characterized.

At the same time that clinical and tenure-track faculty are considering what the EPPs are likely to emerge from their design process, the disruption to the current data system is also of high concern. What data will the designers and faculty want to know about as they move from early design into curriculum and learning experiences that will make up the new “unprograms”? That is a data system design challenge beyond continuous improvement.
A. D1 / D2 Contact Faculty Workforce Student Services Digital (October Major Maps)

B. DESIGN D1 CLINICAL EXP

C. D1 CLINICAL EXP WORKFORCE

**Preformance Assessment**

Building Communities of Education

- D1 Personalized Pathways
- New Roles-Added Value Pathways
- D2 Pathways to graduate degrees/leadership roles

Big “Questions”
- InTASC Standards
- Intrapreneurship
- Design
- Character Ed
- Teaming
- Digital Instruction
- AND Specialists

...and many more on-demand, contextual modules...

**EXPERIENCE**

- VIDEO TC's

**CONTENT**

- Building Communities of Education
- Big “Questions”
- Communities of Educators

**Build a Body of Evidence**

- “Uber” of Personalized Experience
- BUILDING INT'L CONNECTIONS
- GLOBAL DATA METRICS

**“Communities of Educators” Conceptual Models**
- ASU prep/pca
- Buckee
- Avondale K-8
- Miami JP High

**F. ASU PREP WORKFORCE**

**G. WORKFORCE DESIGN FACILITATION? AS NEEDED**

**H. Faculty Research Teams**

Data Strategy Recruiting

**I. “EDUCATION DESIGN CENTER”**

**Figure 2.2 ASU Innovation in teacher preparation**
The Continuous Improvement Process

References


Duty, L., & Kern, T. (2014). *So you think you want to innovate? Emerging lessons and a new tool for state and district leaders working to build a culture of innovation.* Retrieved from https://static1.squarespace.com/static/54e3a3a5e4b0abc9f7b6be/t/56d738929f266eea24ce482/1456945310797/2Rev-TLA_Assessing+Culture+of+Innovation_FINAL.pdf


Part II

From the Researchers
Creating a Data Culture in Educator Preparation
The Role of the States

Michael Allen and Charles Coble

The Emerging State Role

Over the course of many years working inside programs that educate teachers and school leaders as well as working at the interface of educator preparation and state education policy, we have come to recognize several important “truths” about data in the educator preparation context that illustrate both the promise and the challenge of efforts by teacher educators and state officials to use data effectively in the service of preparation program improvement. Teacher educators with whom we have worked range from being skeptical to being keenly interested in using data to inform appraisals of the strengths and weaknesses of program candidates and to foster the improvement of educator preparation programs. This difference in receptivity to using data for improvement is based largely upon legitimate concerns about data’s validity and utility. Rarely is there a shortage of data per se in educator preparation, but what is in short supply are data that are sufficiently valid, reliable, and specific to permit an accurate appraisal of the true strengths and weaknesses of program candidates and completers. Consequently, teacher educators too often make important decisions about candidate and program strengths and weaknesses based upon data that are misleading “noise” instead of upon data that can provide true signals of candidate and program performance.

On the other hand, we have found that teacher educators are often too quick to dismiss some data sources that we believe can be of immense value in appraising program strengths and weaknesses. These include, for example, measures of the academic proficiency of incoming teacher candidates and measures of the effectiveness of recent program completers in improving their P-12 students’ success. Considered in the proper context and alongside other data, such measures can be important indicators of program deficiencies and of relative program strengths, as we discuss later in the chapter.

State officials charged with the responsibility for approving and periodically reviewing educator preparation programs face their own set of challenges. Many states lack adequate staff capacity to carry out thorough periodic reviews of all preparation programs in the state, and consequently
they may rely substantially on the periodic assessments by the national accreditor (currently the Council for the Accreditation of Educator Preparation, or CAEP). National accreditation, however, focuses much more rigorously on the policies, practices, and resources of the provider as a “unit” (education preparation provider or EPP) than on the performance of the provider’s individual certification programs, although many EPPs and states opt for a “review” by outside content experts of the policies and practices of individual licensure field programs as part of the accreditation process (CAEP, 2013). Many states require EPPs to submit an annual or biannual self-assessment that may include evidence of improvements or specific program weaknesses, but some states don’t have the capacity to review these documents carefully. And although states annually collect program-level data in order to comply with Higher Education Act (HEA) Title II reporting requirements, we have perceived a substantial consensus among state officials and teacher educators that Title II data are of little value for assessing program strengths and weaknesses. This is in part due to the sheer volume of the data required to be reported and in part due to the conviction that the data have little value for assessing important program strengths and weaknesses and thus for aiding program improvement (U.S. Department of Education, 2016; see also Crowe, 2010).

We have also perceived an interest among many state officials in offering technical assistance, provided by the state agency, to help strengthen educator preparation programs, especially those that are performing poorly. Even states with regional technical assistance centers, however, lack the staff capacity and expertise to provide sustained technical assistance targeted to problems facing specific EPPs and their programs. Moreover, some state officials have expressed a concern that when an agency charged with accountability for program performance offers technical assistance to lower-performing programs, it creates a potential conflict of interest. This is because offering technical assistance could create a “halo effect” that leads states, perhaps unwittingly, to grade the performance of programs given technical assistance more highly to validate the value of the technical assistance.

The Data on Teacher Preparation Program Outcomes

A fortuitous set of circumstances put us in a position to address these and related challenges directly and systematically beginning several years ago. In 2013, CAEP and Pearson Education commissioned us to provide an assessment of the current prospects for a nationwide evaluation of educator preparation programs. We were also asked to identify gaps in data collection and weaknesses in state and national data systems and to make recommendations to the field on how to improve the quality of data to create a more complete, reliable, and useable data profile of educator preparation.

A major deliverable of the CAEP-Pearson project was a report, Building an Evidence-Based System for Teacher Preparation (Allen, Coble, & Crowe, 2014),
that included a comprehensive framework for analyzing the status of assessment and accountability for educator preparation in the United States based on existing research, an examination of state data and information available in 15 selected states, and exemplary programs or initiatives at national, state, and institutional levels. That framework came to be known as the Key Effectiveness Indicators (KEI), a set of 12 outcomes-focused indicators and 20 measures of preparation program performance in four domains. (See Appendix, Table 1.) We proposed that the indicators could be used for three related purposes: state program accountability, through an annual (or perhaps biannual) review of program performance on the various indicators; dissemination of information to consumers and other stakeholders via an appropriately simplified and (if necessary) redacted published report of program performance; and program self-improvement based on the exploration of the program strengths and weaknesses that might be signaled by the various performance scores.

At the same time as we began to work on the report and develop the Key Effectiveness Indicators, the Council of Chief State School Officers (CCSSO) created a network that eventually involved the participation of 15 states in an effort to implement the policy recommendations of a then-recently released report, Our Responsibility, Our Promise: Transforming Educator Preparation and Entry into the Profession (CCSSO, 2012). The report called for improvements in: (a) state licensure, (b) state educator preparation program approval, and (c) state data collection, analysis, and reporting. The CAEP-Pearson report, and especially the Key Effectiveness Indicators, became important tools for the new CCSSO Network for the Transformation of Educator Preparation (NTEP), providing participating states a kind of benchmark for their progress toward meeting some of the Our Responsibility, Our Promise report’s goals. We served as advisors and technical assistance consultants for the project, and we ultimately expanded our work to over 20 states.

In addition to providing technical assistance, we collected and analyzed accountability processes and measures across the 50 states and Washington, D.C., to produce a report for CCSSO, in conjunction with the NTEP initiative, that attempted to discern the larger picture of teacher preparation program accountability in the U.S. (CCSSO, 2016). The report indicated that roughly half of U.S. states had undertaken efforts to identify appropriate data and measures that could be employed annually to distinguish between high-performing and low-performing teacher preparation programs and thereby motivate state-driven program improvement. Although some states were in the very early stages of the process while others had already begun implementation, most states pursuing this work were on track to activate at least some of their chosen performance measures by late 2017. Specific measures differed from state to state, but the CCSSO report revealed a broad consensus among the states involved in these efforts that the focus should principally be on the results produced by teacher preparation
programs—especially the demonstrated knowledge and teaching skill of candidates and completers—rather than on program elements or structures.

Several other reports on the use of outcomes-oriented data for educator preparation have been released in the last two years. These include accounts of state policies and lessons learned, summaries of how teacher educators are using data for program improvement, the use of data by teacher educators, and research studies critical of efforts to employ outcomes data—in particular, teacher impact data—to evaluate and compare preparation programs (see, for example, CCSSO, 2018; Data Quality Campaign, 2017; Deans for Impact, 2016; Koedel & Parsons, 2014; Mitchel & Aldeman, 2016; von Hippel, Bellows, Osborne, Lincove, & Mills, 2016). Taken together, these reports and the many state and institutional efforts to identify and employ more strategic metrics to evaluate and improve the performance of educator preparation programs indicate that this is presently a robust, important, and growing phenomenon.

Caution, if not skepticism, about the ultimate success of using outcome measures to assess the performance of teacher preparation programs is certainly warranted. Several years ago, partly in response to mandatory measures for educator preparation program evaluation proposed by the U.S. Department of Education, the National Academy of Education (NAEd) undertook an analysis of the kinds of data and measures that had been used or proposed for the evaluation of educator preparation programs (Feuer, Floden, Chudowsky, & Ahn, 2013). The analysis concluded that there was an inadequate research base to confirm that any of these measures yielded a valid and reliable assessment of program strengths and weaknesses. Instead, the NAEd report recommended that state officials and others seeking to use such measures for program accountability or system improvement should address seven questions to inform the process as they move forward. That may be sage advice, but the reality is that approximately half of the states in the U.S. already have developed and implemented annual program performance measures. Our experience working directly with most of these states leads us to believe strongly that the development of performance accountability systems is worth pursuing despite potential pitfalls and limitations of the enterprise.

Our work with the states has focused primarily on helping them advance their thinking and understanding concerning (a) the appropriate indicators to determine program performance, (b) the relative information value and reliability of various kinds of measures, (c) the data requirements for moving forward, (d) effective and defensible accountability policies and practices, and (e) the complementary responsibilities of state agencies and EPPs in the program improvement process. We share here a summary of what we have observed and of what we have come to perceive as the challenges and promise of the efforts states are undertaking. And we suggest some prescriptions for states and their stakeholders as they move forward.
Creating a Data Culture

Generating the Program Performance Data

The actual indicators and measures that states have adopted to generate the performance data for program review and accountability differ substantially between the states in overall scope and in the specific measures themselves. Virtually all of the 20-plus states that are constructing new accountability systems, however, are committed—at least in theory—to building a system that is tied to program outcomes, including the performance of program candidates and completers.

The conceptual leap involved in using indicators of candidate or completer outcomes to assess the performance of teacher preparation programs is the assumption that programs have an impact on the outcomes measured under the indicators. That conceptual leap was first taken in 2003 by officials in the state of Louisiana who sought to discover whether student growth could be linked to individual teacher education programs through the state’s new value-added assessment model (Noell & Burns, 2006). Their research found that there were, indeed, statistically significant differences in measured performance outcomes among programs at the very high and very low ends of the scoring distribution. Follow-up efforts identified potential weaknesses in low-scoring programs that might account for the low score results. The state has continued to use program-level value-added assessment results, as well as other indicators of program proficiency, to investigate deficiencies in particularly low-scoring programs and implement corrective measures. In one instance, for example, a state-initiated investigation found that a program whose completers had particularly low value-added scores in reading lacked adequate faculty capacity in reading education (Gansle et al., 2015). As other states with value-added systems began to develop similar efforts to connect teaching outcomes to teacher preparation programs, the Obama administration—through its Race to the Top initiative—pushed grantee states to develop teacher preparation program accountability systems based on teacher impact and other strategic outcomes (Crowe, 2010).

Many individuals in the teacher education world have regarded this push as a “leap too far” (Indiana University School of Education Value-Added Task Force, 2011), but the connection between teacher outcomes and certification programs has antecedents and corroboration in the empirical research. Numerous studies have confirmed that different kinds of teacher preparation programs have variable impacts on completer outcomes, such as teacher value-added measures, although in some cases the variations in impacts for completers within the same program are as great as the variations in impacts between completers of different programs (Boyd, Grossman, Hammerness, et al., 2008; Koedel, Parsons, Podgursky, & Ehlert, 2015).

To be sure, the impact of preparation programs on some measures, such as candidate retention in the profession, seems more difficult to isolate. And since candidates’ knowledge of their teaching subject is gained in the arts
and sciences and, thus, largely outside of schools of education, preparation program accountability for this also seems more remote. Nevertheless, we believe (and states and CAEP believe) that preparation programs do have responsibility for such outcomes and that program performance data collected on the assumption of such connections can provide important information to teacher educators in their efforts to improve the effectiveness of their programs. Such data may identify a common weakness in program completers’ content knowledge, for example, that could lead to appropriate changes in content courses. Similarly, data on candidate retention—regardless of responsibility—aids state officials in their efforts to ensure that programs in the state can meet the identified state needs for teachers.

Program Performance Indicators

The development of a statewide, indicators-based preparation program improvement and accountability system is a long-term process. For both strategic and practical reasons, states need to implement indicators and accountability processes gradually over time. Education officials must select indicators and measures that are applicable to all programs, regardless of configuration. They must test the validity and reliability of their measures and their efficacy for program improvement. They must identify which program performance levels on the various measures are acceptable and which are legitimate cause for concern. And they must determine whether the data are adequate to assess reliably the outcomes they are intended to measure.

Virtually all the indicators states employ, however, will fall within the four focus areas we identify in the Key Effectiveness Indicators: (a) Candidate Selection and Completion; (b) Knowledge and Skills for Teaching; (c) Performance as Classroom Teachers; and (d) Contribution to State Workforce Needs. Development of the KEI was motivated by what we perceived as states’ desire to craft a limited set of essential indicators—and ultimately measures—that would (a) be useful for program improvement, (b) be equally applicable to all programs in a state (traditional and nontraditional) without prescribing program content and structure, (c) be focused principally on program outcomes rather than inputs, (d) be compelling and transparent to a variety of stakeholders, and (e) respond to the concerns that both experts and the greater public have expressed about the effectiveness of programs and the caliber of individuals entering the teaching profession. Eventually we framed the four categories (or domains) of program assessment data, 12 indicators, and suggested measures for each of the KEI indicators.

We have found considerable receptivity to the KEI among the states with which we have worked, primarily because the KEI provide an intuitive framework for understanding the landscape of indicators and performance measures in teacher preparation. And although not all states and programs intend to adopt the KEI in their entirety, many use it as a check against the adequacy of the indicators and measures they currently use or proposed to use for their accountability systems. To facilitate this process, we developed
A Guide to the Key Effectiveness Indicators, which provides a good deal of context and discussion about measuring preparation program performance and accountability (Teacher Preparation Analytics, 2016).

The KEI indicators and measures are intended to serve as a “first-line” assessment of program performance and to meet the need for the kind of “data dashboard” that some states have developed for both K–12 schools and post-secondary institutions. Additional information about a program not captured by the KEI is necessary to identify how specific teacher preparation program coursework, program design and operation (e.g., partnerships and clinical experiences), and feedback processes may produce the outcomes that yield the KEI measures.

We summarize the individual KEI indicators and the research that supports their inclusion below, grouped into the four program performance domains.

Candidate Selection and Completion

Each indicator in this domain of the KEI is relevant and valuable to an overall assessment of program effectiveness. These indicators may appear to be “inputs” rather than “outcomes,” but the measures for each indicator reflect the outcome of deliberate processes—candidate selection, assessment, and (to the extent offered) program support—that a program or EPP has in place. The indicators address vital concerns that teacher educators, policy makers, education leaders, and the wider public have about the academic strength, diversity, and fit for teaching of the candidates who enter and complete teacher preparation programs. The judicious selection of teacher candidates should increase the likelihood of their success in the program, effectiveness in the classroom, and long-term commitment to the teaching profession. (Goldhaber & Walch, 2013).

Academic Strength

Academic strength has become one of the most controversial issues in discussions about teacher preparation program quality because of concerns that increasing academic proficiency requirements for program entry will seriously reduce the available pool of minority and/or low-income teacher education candidates, whose incoming academic proficiency may have been compromised by an inferior elementary and secondary school education. Defenders of the importance of high academic standards for program admission cite numerous studies that have found or suggested a correlation between the academic proficiency of teachers and their success in the classroom. A report on teacher preparation by the National Research Council (2010), for example, found some evidence for the positive impact on outcomes of program selectivity, citing studies conducted in Florida, New York, and North Carolina. In the international context, a McKinsey & Company report found that the highest performing national school systems in the world recruited teachers primarily from the top third of the university
graduating class, while only 23% of U.S. teachers came from the top third (Auguste, Kihn, & Miller, 2010). In addition, a 2016 study we conducted for CAEP found little support for a direct correlation between SAT scores and teacher effectiveness but significant support for correlations between SAT scores, college completion, and content knowledge proficiency, and additional support for the correlation between academic proficiency generally and teacher effectiveness (Coble, Crowe, & Allen, 2016).

The inclusion of measures of academic strength in a preparation program profile is not solely a matter of its relationship to teaching effectiveness, however. It also can provide assurance to concerned policy makers and other stakeholders that the candidates entering (or completing) teacher preparation programs are on par academically with individuals entering other learned professions. In turn, evidence that new teachers are indeed as capable academically as those in other fields raises the prestige and respectability of the teaching profession.

Teaching Promise

Preparation programs, school districts, and national organizations like Teach for America and UT all seek to measure individual attitudes, values, and behaviors that may predict suitability for and success in teaching. The research evidence linking specific beliefs, values, or habits to teacher effectiveness is not robust, but it is sufficient to suggest the value of screening applicants to teacher preparation as is done routinely in other professional fields and employment recruitment.

While it is not difficult to imagine EPPs screening applicants with an instrument such as the Duckworth team’s Grit Scale (Duckworth, Peterson, Matthews, & Kelly, 2007), it is harder to envision reporting results for individual candidates or for cohorts of applicants and admitted students in a way that supports easy-to-use comparisons across programs or states. This is one of two current limitations of this indicator as a measure of program effectiveness. The second is the need to find one or more “teaching promise” metrics that can be linked directly to important components of high-quality classroom teaching and that might be used as the basis for comparison between programs of incoming candidates’ promise as teachers. Working with NCS Pearson, the state of Missouri has developed an assessment that employs such metrics—the Missouri Educator Profile—but currently it is used only for candidate development, and scores are not reported (NCS Pearson, 2013).

Candidate and Completer Diversity

Policy leaders and teacher educators support the idea that the teaching force should be diverse, not only to provide opportunities for talented individuals but also because of the increasing diversity of the K–12 student population in the United States. Currently, about 82% of U.S. K–12 teachers are white, 7% are African-American, and almost 8% are Hispanic. Men comprise
24% of the K–12 teaching population (National Center for Education Statistics (NCES, 2014). The demographic composition of the K–12 student population is far more diverse than that of the teacher workforce.

Most preparation programs collect information about the demographic composition of applicants, admitted students, and program completers, though little of this is widely shared outside the program. While not enough is yet known about the empirical relationships between teacher demographics and K–12 student outcomes, the demographic composition of program candidates and completers is a policy concern in almost every state. If the goal is to ensure that programs are indeed increasing the diversity of the teacher workforce, then it is particularly important to collect, report, and analyze comparable data on the diversity of admitted candidates and completing candidates from an admitted EPP cohort. To ensure that admitting a diverse pool of candidates is more than an exercise in affirmative action, we believe that EPPs should be held accountable for ensuring that the candidates admitted, regardless of socioeconomic, ethnic, and racial background, have an equal chance of successfully completing the program. For programs that admit high numbers of minority and low-income candidates—often first-generation college students who may have weak academic backgrounds and low self-confidence as learners—an increased investment in academic remediation and other support services may be required to increase their program success and completion.

**Knowledge and Skills for Teaching**

The indicators in this category are the most immediate outcomes of pre-service preparation, with fewer covariates in play than in the connection of postgraduation completer outcomes to preparation program variables. They also are the most actionable in terms of programs’ ability to address candidates’ low performance. Measures of knowledge and skills for teaching also provide a check on the measures of completers’ classroom performance; it would be very unlikely, for example, to see high scores on classroom performance and low scores on teaching knowledge and skill. Assessments for these indicators are commonly required for teacher licensure, although some may not be sufficiently robust to measure true mastery of the knowledge and skills in question.

**Mastery of Teaching Subjects**

While it seems self-evident that mastery of content is a prerequisite to effective teaching in the subject matter one is to teach, the empirical evidence available to support the link between adequate content knowledge and effective teaching can be confounded by at least three factors:

1. Tests currently in use may not assess the kinds of content knowledge that are tied to important teaching outcomes.
2. The standard for “adequate” content knowledge is set by states, and the threshold is often a function of supply-demand considerations unrelated to actual teaching ability or effectiveness. Thus, states may set a low passing threshold or may reduce course requirements in a given field (often the sciences) to ensure adequate supply of teachers.

3. Unlike other professions, virtually every state uses its own set of teacher knowledge and skills tests; if two or more states use the same test(s), passing scores are often set at different points.

Nonetheless, there is strong research in certain content areas, such as reading and mathematics, that clearly articulates what teachers need to know for their students to learn and grow academically. In reading instruction, for example, the work of the National Reading Panel (2010) identifies the essential components of scientifically based reading instruction as well as the implications of high-quality research about reading for school curriculum, teaching, and teacher preparation.

Similarly, William Schmidt and his colleagues cite studies demonstrating that (a) teacher knowledge of mathematics “has a significant relationship to student learning” and (b) teacher preparation programs influence mathematical content knowledge of teachers and that an important relationship exists for P-12 students between their exposure to mathematical content knowledge and mathematics learning (Schmidt, Burroughs, Cogan, & Houang, 2017).

Subject-Specific Pedagogical Knowledge

The knowledge of how best to teach the content and the common conceptual challenges to learning it is undervalued, under-assessed, and little understood. Work by Deborah Ball and her colleagues corroborates its importance (Ball, Thames, & Phelps, 2008). A partnership between ETS and TeachingWorks, led by Ball at the University of Michigan, has produced the National Observational Teaching Examination (NOTE) series of license tests (ETS, 2011). These include a Content Knowledge for Teaching (CKT) assessment that is designed to assess candidates’ grasp of subject-specific pedagogical knowledge more directly and comprehensively than has been possible until now. Moreover, since subscales within CKT address fundamental content knowledge in a tested teaching field, ETS believes that CKT also may provide an adequate assessment of candidates’ general grasp of their teaching subject.

Several states now require the administration of either the edTPA or the Praxis Performance Assessment for Teachers (PPAT) as a summative program assessment or a requirement for teacher licensure (see ETS, 2018; Stanford Center for Assessment, Learning, and Equity (SCALE), 2014). These assessments test candidates’ pedagogical subject knowledge specific to their teaching field, but only a very narrow slice of it.
Creating a Data Culture

General Teaching Skill

The classroom teaching performance of candidates is a fundamental “within program” outcome measure of program quality and core competency required for teacher licensure. Until recently, the most commonly employed assessment of a candidate’s teaching skill was the Praxis Principles of Learning and Teaching (PLT), a paper and pencil assessment. With the development and increasingly wide adoption of the edTPA (and the PPAT), assessment of candidates’ teaching skill has become far more robust and “authentic” (Peck, Singer-Gabella, Sloan, & Lin, 2014). Both assessments use portfolios completed by teaching candidates during their student teaching experience which typically include video clips of instruction, lesson plans, samples of student work and teacher feedback, and candidates’ reflective commentaries. Teacher preparation programs can use these data to assess candidates’ readiness to enter the teaching profession and as a source of evidence for programmatic reforms. While voluntary in some states, other states require candidates to pass the edTPA, PPAT, or comparable state-developed assessment as a condition for teacher licensure (Lys, L’Esperance, Dobson, & Bullock, 2014).

Completer Rating of Preparation Program

Program completer ratings of satisfaction with teacher preparation programs are being used by a growing number of programs, states, and university systems for assessing program quality and for program improvement. The American Psychological Association Task Force on teacher preparation program improvement and accountability concluded that “surveys can be very useful as a program evaluation tool with former teacher candidates within a year of graduation and several years after graduation” (Worrell et al., 2014). Most states that employ such a survey administer them to program completers prior to graduation when they can ensure a high response rate. Some states, however, administer them well into teachers’ first year of teaching (and, in at least one state, also the second year), but the response rate is much lower because of attrition, lack of contact information, and absence of leverage to compel responses. The justification for a later administration of the survey is the supposition that teachers who have a year or more of actual full-time teaching experience have a more helpful perspective on the adequacy of their preparation program than do teachers who have yet to begin their career.

We believe that such a survey indeed has the potential to be an important vehicle for program evaluation and improvement. To serve that function, however, the survey instrument needs to be constructed and scored so that it promotes real differentiation in respondents’ assessment of different aspects of their program. Not only is that presently not always the case, but survey results often show little difference in scores between programs. That may
simply reflect the limitations of this vehicle for program accountability, but it may also point to the need to redesign either the instruments or the way they are scored. Moreover, some states and EPPs have used these sorts of surveys more informally for quite some time—before they were intended to be the basis for program accountability—and earlier surveys, if unrevised, may lack the rigor and validity necessary to serve an accountability function that has potential consequences for EPPs and their programs.

**Performance as Classroom Teachers**

Indicators in this category of the KEI—especially teacher impact on K–12 students—are what people generally regard as the most compelling outcomes for assessing the performance of teacher preparation programs. It seems reasonable to expect that teachers prepared in effective programs should perform better in the classroom, other things being equal, than those from weaker programs.

Many experts are concerned that in-school factors that can affect a teacher’s performance call into question the validity of such in-classroom assessments. However, the validity of the three indicators we suggest in the KEI, at least for the specific instruments employed, was corroborated in the Gates Foundation’s Measures of Effective Teaching (MET) study (Cantrell & Kane, 2013). The study also found that the use of any two or all three of the indicators increased their reliability as measures of teacher effectiveness; teachers who scored high or low on one indicator tended to score high or low on the others.

The most frequently used of the three indicators is an observation-based assessment of Demonstrated Teaching Skill (CCSSO, 2016). Fewer states use value-added or student growth data in their assessment of programs, and still fewer solicit K–12 students’ perceptions of their teachers’ effectiveness, although we have noted substantial interest in adopting the latter as we have worked with states over the past few years.

The value of these indicators as measures of program performance also depends upon the validity and reliability of the specific instruments and assessment methods employed. It is important to note, however, that the instability of any particular measure in assessing individual teacher performance is greatly reduced when multiple individual measures are aggregated into a measure of a whole teacher cohort—and even more so the larger that cohort is.

**Teacher Impact on K–12 Students**

More than any other single indicator, teacher impact on student learning is the “holy grail” for assessing the effectiveness of teacher preparation programs. Nationally, we found that 13 states have adopted this indicator (CCSSO, 2016). Some states using it as an indicator of preparation program
performance have found, however, that, as with other measures, it does not differentiate substantially between the performance of most programs and reveals as much variation in completer performance within individual programs as between them (Koedel et al., 2015; von Hippel et al., 2016). Moreover, researchers disagree about the relative validity and reliability of various kinds of impact measures (e.g., value-added vs. student growth) (Ehlert, Koedel, Parsons, & Podgursky, 2013; Goldhaber, Walch, & Gabele, 2013). Nevertheless, researchers continue confidently to employ value-added and other impact measures in their studies (Goldhaber & Little, 2012; Henry et al., 2010).

Teacher impact is potentially the strongest anchor for the assessment of preparation program performance—the one measure that is not a just a proxy for effective teaching. And as we noted earlier (p. 8), states that have compared very high and very low program scores on teacher impact measures have found important differences in the structure of those programs. Based upon our conversations with officials in preparing the CCSSO 50-state report, however, we perceived a backlash in several states attending the use of teacher impact measures—a trend that may be partly fueled by relaxed rules on teacher evaluation under the Every Student Succeeds Act (CCSSO, 2016; Sawchuck, 2016, January 6). This often involves opposition from teachers and their unions—in some cases to any use of value-added (or teacher impact) data in teacher evaluations, and in others to the use of teacher-impact data for any purpose other than confidential teacher evaluation (Croft & Buddin, 2014). This trend could portend a growing moratorium on states’ continued use or adoption of the indicator in teacher evaluations as well as preparation program evaluations. Also, because the states have greater policy autonomy under ESSA, and because no incentives or requirements are specifically tied to states’ implementation of teacher impact measures, we believe it has become less likely that states will build value-added or other sophisticated data systems that measure teacher impact because of the controversy and the costs involved (Loewus, 2017, November 14).

**Demonstrated Teaching Skill**

CCSSO’s 50-state analysis revealed that 10 states use observation-based assessments of new teachers as an indicator of their preparation programs’ program performance (CCSSO, 2016). Many districts undertake such an assessment as part of induction and mentoring programs and their own evaluation of novice teachers. We believe this is a far more valuable and reliable indicator of new teacher skill than the surveys many states employ that ask principals and supervisors to assess the preparedness of new teachers. Moreover, classroom observations of teachers may be more feasible to implement and more readily acceptable politically for many states than teacher impact measures. For use as a state program performance measure,
Michael Allen and Charles Coble

However, it is critically important that the observation assessment be rigorous and valid, that it be administered by trained observers who ideally have expertise in the subject of the assessed teacher, and that standards and measures employed are compatible between districts. Fortunately, a growing number of quality classroom observation instruments are available. For example, the Danielson Framework for Teaching, Teachstone’s Classroom Assessment Scoring System (CLASS), and several others used in the MET project have significant support in research (Cantrell & Kane, 2013; Danielson Group, 2017; Teachstone, 2018).

**K–12 Student Perceptions**

Our work with states has revealed growing support for the notion that K–12 pupils’ assessments of their teachers’ effectiveness can provide valuable data for teacher improvement. The MET project supports this idea in its finding that pupil observations of their teachers are well-correlated with value-added and classroom observation measures (Cantrell & Kane, 2013). In addition, an American Psychological Association (APA) task force report states that appropriately constructed student surveys of their teachers are more highly correlated with student achievement than either teacher self-evaluation or principal ratings (Worrell et al., 2014). While this indicator is largely aspirational, our work with CCSSO indicated that several states are moving toward its near-term adoption (CCSSO, 2016).

**Contribution to State Teacher Needs**

This set of indicators is important to state and district officials in their efforts to know how well different preparation programs in the state are responding to the need for teachers around the state, especially in subjects and districts with serious teacher shortages. EPPs have different missions and student populations, and some institutions historically enroll many out-of-state students who return to their native states upon graduation. But for state-supported institutions in particular it seems reasonable to expect most EPPs to respond to state teacher needs, to trim programs that regularly produce teachers in areas of large surplus, and to enlarge programs that produce teachers in shortage subjects and for hard-to-staff schools.

**Entry and Persistence in Teaching**

The supply and demand picture for teachers in the U.S. historically has varied greatly by teaching field and geographic location (Yasin, 1999). Although state and federal supply and demand forecasts for the profession likely do have some impact on the priorities of education preparation providers, the production system in the U.S. is inherently inefficient. Especially in traditional undergraduate teacher education programs, it is the supply...
side—students’ interest in the teaching profession and the willingness of programs to accommodate them—that has the most influence on how many teachers are produced and in what fields. Some programs are the exception, especially those that work in close partnership with specific districts to train and place teachers. But most other programs play a much more limited role in ensuring that their graduates obtain jobs. To be sure, preparation programs are not solely responsible for job placement. However, the fact that some programs have much higher placement rates than others—whether through better initial screening of candidates, stronger relationships with districts, or greater efforts to tailor supply to demand—is sufficient justification for attributing some responsibility to programs for their placement rates and finding out whether practices of programs with higher rates can be adopted by programs that have lower rates.

Similarly, preparation programs are not solely responsible for the persistence or turnover of their completers once they enter teaching. Given some of the causes and consequences of teacher turnover, however, persistence in teaching is a program outcome that can help to align the interests of producers and employers (Henry, Fortner, & Bastian, 2012). It may be that, as with the new teacher residency programs, all programs might need to restructure their priorities to provide support for their recent graduates in the field (NCATE Blue Ribbon Panel, 2010; National Center for Teacher Residencies, 2017).

It is difficult, of course, to track program completers once they leave their preparation programs, and this poses a challenge not only for measuring completer placement and retention rates but also for measuring teacher performance and impact. Obtaining reliable data on persistence as a program outcome requires data systems that enable all programs to locate completers in the schools and districts where they teach, and this likely needs to be a state-level commitment. Even then it will be a much greater challenge to track program completers who teach in private schools or out of state.

**Placement/Persistence in High-Need Schools/Subjects**

Persistence rates in key subject areas and in high-need/high-poverty schools are also important to track, report, and analyze. The highest turnover rates are in low-performing and high-minority population schools (SERVE Center, 2006; Clotfelter, Ladd, & Vigdor, 2007a). Programs, schools, and policymakers need stronger incentives to address this problem more aggressively; public reporting of these rates as a program performance measure will help. And although here, too, preparation programs have only limited responsibility for how long their graduates stay in challenging schools, significant variations between programs in the persistence rates of their graduates in low-performing and/or high-minority schools may point to differences in program practices that account for at least some of the variations in persistence outcomes.
Program Performance Measures

Ultimately, indicators are only as useful as the quality of their measures allows them to be. For the KEI system we propose 20 measures to employ under the 12 indicators. (See Appendix, Table 1.) These are not the only possible measures that might be used, but these measures have significant informational advantages over many more commonly employed measures and are better aligned to the relevant empirical research in the field. One change we propose, for example, is for states to move from using “pass rates” on assessments as a measure of program performance (a vestige of federal reporting requirements still enforced under Title II of HEA) and instead employ measures such as median candidate or completer score. Although the research does not support a correlation between passing licensure examinations and teacher effectiveness, research does support a connection between actual examination score (raw score or standardized score) and teaching success (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009; Clotfelter, Ladd, & Vigdor, 2007b; Coble et al., 2016).

We also propose that, insofar as possible, measures reflect scoring distribution. We suggest measuring the number and percentage of candidates scoring in the upper and lower third (tercile) or quartile in a statewide or national distribution of assessment scores. Mean scores alone mask variation in scores, and that could hide the fact that a program with a cohort mean slightly above the state average on an assessment may nevertheless have a large percentage of program candidates or completers with weak performance.

Finally, we believe that measures used must, whenever possible, reflect individual certification program-level outcomes rather than EPP-level aggregate outcomes. Measures at the provider level yield misleading information that, once again, can mask significant variations in individual program performance. And for program improvement purposes, it is only data at the program level or very close to it (e.g., certification track such as elementary education or secondary mathematics that could be served by several different programs) that signal specific program strengths and weaknesses and can guide diagnosis.

Implementing the System: Challenges and Lessons Learned

For a variety of reasons many states have begun to place more emphasis on annual program reviews and reports. States commonly review educator preparation programs for reapproval every 5 to 10 years (depending upon state statutes). Many states grant reapproval status to programs in conjunction with national (CAEP) accreditation of the EPP, a process of program review by outside teacher educators that may or may not include a review of individual certification programs by either a national disciplinary association (e.g., the National Science Teachers Association) or as part of an EPP
Creating a Data Culture

self-study submitted as evidence for accreditation. A few states require all their EPPs to be nationally accredited, more require all public EPPs to be accredited, but most states leave that decision to individual EPPs themselves. States also commonly require an annual or biennial report on certification programs, particularly if there is past evidence of program insufficiencies. CAEP also requires annual reports, but only at the EPP level (CAEP, 2013). A few states have no periodic program reapproval process once a program has been initially approved, and these states rely solely on annual reports as an assurance of continued program quality.

Thus there is a patchwork of quality assurance policies in educator preparation among the states, and even a patchwork within states that offer both accreditation and non-accreditation paths to program reapproval. From a state accountability standpoint, a dual system of program approval is highly problematic; it means that programs do not necessarily have to meet the same standards to be state-approved. Moreover, some states—including almost all states that participated in CCSSO’s NTEP network—have recognized that a multiyear reapproval process misses the opportunity to address program deficiencies more promptly and that more frequent review promotes a continuous improvement process that could benefit all programs. These states and others have decided to invest considerable effort in creating an annual process of review for EPPs and their individual certification programs.

The new program accountability processes that states are developing hold out the promise of a more transparent and—in the eyes of a public suspicious of peer review—less self-interested preparation program assessment. Focused on outcomes rather than inputs, the reports generated by these systems may be more compelling to stakeholders who are interested in outcomes and who lack the professional understanding of program processes. Moreover, as annual events the new state reports may convey a greater sense of urgency among preparation programs to address their failings and boost public confidence in their commitment to do so.

The states that we have worked with in developing systems of annual program review are at various stages in system development. Some states have been refining their systems for many years—in some cases proceeding rapidly and in other cases in fits and starts. Some states are just beginning to take on such a project and are making discernible—if still tentative—progress.

Key Challenges

The challenges to developing state accountability systems are formidable. On the technical side, they involve all the issues related to data summarized above as well as the need to monitor the system and its impact and, if called for, make system improvements to make necessary changes. On the political side, states must ensure that the indicators and measures are meaningful and satisfying to a variety of constituents who have a stake in the performance
of preparation programs. Education leaders must gain broad support of these stakeholders for the indicators and measures they choose—a tall order when indicators like “teacher impact” (regardless of how measured) can be a lightning rod for opposition. And states must design and implement an accountability system that seems useful and fair—above all, to the programs it is intended to serve—and that motivates program administrators to make improvements while maintaining their willing participation. The many challenges are not easy to overcome, and we discuss some of them in more detail below.

Inevitability of Imperfection

State officials and teacher educators with whom we have worked in states already heavily engaged in the development of these program review and accountability systems overwhelmingly approach their assignment with admirable competency, dedication, and thoughtfulness (see, e.g., Delaware Department of Education, 2015; Hibpshman, 2013). Most state officials are completely aware of the limitations of the enterprise—especially less-than-ideal data and politically and methodologically fraught indicators and measures. They know that their task is to create and support a system that not only must make compromises and has few proven models to draw upon but also, partly because of those limitations, has many skeptics and detractors. We have been impressed that they move forward, nevertheless, out of the conviction that whatever the limitations of the system under development, it is still likely to be an important advance over the status quo and can be honed and improved over time.

Risk of Marginalization

An inherent challenge to the development and implementation of effective program review and accountability systems is that, in almost all states, the new outcomes-oriented systems must operate side by side with more traditional inputs-focused requirements for accountability—those, for example, in periodic accreditation or state approval reviews of EPPs. The ability of the annual outcomes measures to drive program quality can be marginalized when a more familiar program review system exists alongside it, especially if the incumbent system is the decisive factor in programs’ continued permission to operate. Thus, unless the annual review systems are invested with power to drive program change by themselves, they risk being effectively superfluous.

Diversity of Program Characteristics

Another challenge states face in developing truly comprehensive systems is to capture outcomes of preparation programs with widely varying design features and operating characteristics. States must ensure that the outcomes
they select—and employ for program accountability purposes—are equally appropriate and equally useful to the different kinds of programs that exist. This includes ensuring that any measures, benchmarks, or rubrics used to evaluate program performance capture reasonable expectations and standards for all programs. This does not mean that all programs can be expected to perform equally well on all indicators, because programs do have differences in mission that may affect who enrolls or the kinds of teaching positions completers seek. But these exceptions should be relatively few, and states—and EPPs—must aspire to improve their performance even on indicators in which they are inherently disadvantaged. Programs, for example, whose completers have comparatively high attrition out of the profession—presumably because they teach in more challenging schools—should be expected to work with their completers’ districts to tackle the causes of attrition rather than simply accept it as an inevitability.

Likewise, differences in institutional and program size also may call for some accommodations in a statewide accountability system. Small institutions are likely to have more limited staff available to collect program data and fewer resources to execute program improvement plans that respond to deficient performance scores. And small programs also are more likely to face a “small-n problem”—i.e., have so few candidates or completers with data on some measures that the validity and reliability of the measures is called into question. This could be handled by aggregating the data on candidates over several years and cohorts, but then the scores would be based on some data that are relatively old and could well mask significant changes in the newest cohort scores that might reflect important changes in program operation. Data privacy is also a concern when n-size is very small, because it becomes easier to associate performance scores with known candidates or completers (see Deans for Impact, 2016).

Political Impacts

Still another challenge is the “mixed blessing” that states receive when the impetus for the development of an annual outcomes-based program accountability system comes from a legislative mandate. When such a system is written into state law, it carries significant weight and authority. However, legislation can sometimes prescribe system features that are ultimately not defensible methodologically or financially, or that derive from an ideological agenda that prescribes dramatic but simplistic policies instead of more thoughtful and potentially effective ones. More than one state we have worked with has spent considerable time developing indicators, measures, and related accountability policies to satisfy a legislative mandate only to find that mandate overturned or amended and the indicators and policies now moot. The more states can bring all stakeholders to the table at the beginning stages of developing preparation program review systems, the more likely that progress will be sustained.
Finally, not all states have the resources to develop sophisticated program performance accountability systems. Small states with limited financial resources not only may be unable to build the kind of database these systems require but also may have too few staff to carry out the implementation and continued administration necessary for the systems to be effective. In states with a very small number of preparation programs, the cost of implementing a program accountability system that relies on a sophisticated database may not be justified by the benefits.5

States that have limited resources to develop operational program review and accountability systems also may not have adequate resources to sustain them—at least at the level necessary to ensure the system’s effectiveness and reliability. It will be instructive, for example, to see if states that were able to count on Race to the Top funding for the initial development of their preparation program review systems can maintain them at their current level of operation, let alone expand them to include other performance measures, now that the support has ended. Funding issues also threaten the possibility of making continuous improvements to these systems to increase data capacity, adopt better assessments, and ensure adequate training for the various kinds of evaluators the system will require (e.g., to observe teachers’ classroom performance).

Key Lessons Learned
Education leaders in states that have been engaged for considerable time in the development of new outcomes-oriented program review and accountability systems likely could each write a book about the lessons they learned in the process. Indeed, several recent publications, some connected with the NTEP initiative, attempt to draw just such lessons (see, for example, CCSSO, 2016; CCSSO, 2018; Mitchel & Aldeman, 2016). Their conclusions are consistent with the most important lessons we have discerned from our work, which we share immediately below.

Not all Data and Measures Are Equal
Several lessons relate to the performance measures states use. One is that not all data and measures are equally useful and equally actionable. We noted earlier that some measures, such as assessment pass rates, are information poor. And although measures that give average percentile rank are generally a step forward, they should be supplemented by measures of score distribution to give the full performance picture. States also fall short when they provide only EPP-level measures. For measures to have any value for program improvement, in almost all cases they need to be measures of the performance of individual certification programs. As we noted earlier, the
aggregate EPP-level measures that states often construct not only fail to pro-
vide guidance for improvement, but they are misleading averages that mask
potentially important variations in the performance of individual programs.

Not All Measures Differentiate

A related lesson that several states have learned is that preparation program
performance accountability systems may not always provide the degree of
differentiation in performance that was expected. This may be because pro-
gram performance on many of the indicators is relatively similar. It may be
a function of measures that are information poor, based on permissive stan-
dards of performance (often the case with supervisors’ assessments), or not
sufficiently sensitive to pick up important differences in performance (e.g.,
when collapsing widely varying scores into a 1–4 rating). But in some cases,
it may reflect the significant differences in the performance of candidates
and completers within all programs. This is consistent with the finding of
several research studies that conclude that, overall, there can be as much
difference in performance within programs as there is between them (Boyd,
Grossman, Hammerness, et al., 2008; Koedel et al., 2015; von Hippel et al.,
2016). If the nation’s goal for teacher preparation is for all program com-
pleters to be capable teachers from day one (CCSSO, 2012), reporting those
within-program differences yields data that are every bit as important as
data comparing different programs.

A Pilot Phase Is Critical

Because of the difficulty and uncertainty of this work, states have learned
that a pilot phase of program implementation is extremely important, both
for determining the measures that will be most effective and for gaining
the buy-in and trust of the stakeholders involved. The pilot phase is an
opportunity, through trial and error, to validate the indicators, determine
the quality and availability of required data, and calibrate the measures to
determine which scores are likely indicative of satisfactory, exemplary, and
problematic program performance. This is also a time when institutions and
states troubleshoot their capacity to collect and manage data for reporting
purposes. Because of the complexity and sensitivity of the accountability
system development process, almost all states we have worked with strategi-
cally incorporate one or more pilot phases into their efforts.

Accountability Is Fraught

In our experience working with states, the most sensitive and controversial
aspect of the system development process is the policies and practices states
employ to make their indicators and measures serve program accountabil-
ity. To a large extent, this is a question of the specific consequences imposed
upon programs that perform poorly on the selected measures. But it’s also a question of how (and whether) program performance is reported to the public and whether programs perceive the accountability system as a helpful stimulus for program improvement or as a punitive hammer.

A few states have issued or continue to issue an annual program performance report that is intended for public and state agency information only and not as a direct accountability vehicle. In these states, programs that score poorly on various performance measures may be shamed or subject to criticism from the public, advocacy groups, and politicians, but no specific state agency action is taken on programs’ performance scores. At most, state officials in these states may use the history of program performance on annual reports as part of the evidence for multiyear program approval decisions.

All the states that we have worked with, however, have (or plan to have) annual performance reports that have direct, state agency-initiated accountability consequences. In some states, the annual program performance reports are not publicly available. In others, select performance data or summarized data on program performance eventually become public information (often after initial review by the programs)—sometimes conveyed in report form and sometimes accessible through a database. Some states issue a detailed performance report to the EPPs and their programs and issue a less detailed report for the public.

The consideration that appears to have the greatest consequences for the long-term efficacy of the program accountability process is whether the system promotes—and is perceived by teacher educators to promote—a cooperative state agency/EPP process of program improvement. Systems perceived by EPPs as largely punitive will face ongoing resistance and will likely be ineffective over the long term. This does not mean that states should refrain from requiring EPP compliance and EPP action to address poor performance; indeed, it is an obligation of the state to be the guardian of quality control. But it means that states need to gain as much buy-in as possible from teacher educators and work jointly with them and other stakeholders both to gain acceptance of the process and to ensure that it does indeed support and promote program improvement.

A few states weight the indicators they apply to derive a composite program performance score. Programs with higher composite performance scores are deemed satisfactory (or, potentially, outstanding), and programs with lower composite scores may be deemed in need of intervention. This is the approach to preparation program accountability that was required in the proposed new HEA Title II regulations concerning annual programs reporting (U.S. Government, 2016).

Other states have adopted an approach that establishes benchmarks for each program performance indicator, and when a program fails to meet the benchmark for satisfactory performance—especially in multiple years—it triggers an inquiry from the state agency to determine whether the low
performance score does indeed signify a serious problem requiring remediation. States that follow this approach assign no composite score to programs, but programs with a comparatively high number of low scores on performance measures are likely to prompt greater concern and more aggressive intervention. And programs with a comparatively greater number of high scores or an extremely high score on a specific indicator hopefully will prompt inquiry as potential exemplars of program practice.

The composite score approach to the program performance indicators has several advantages over the indicator-by-indicator approach. A single overall score carries a lot of heft and so is likely to be a much stronger driver of program response. State agency staff intervention is likely to be minimal in this approach because only programs with low overall scores will be the focus of attention. And the public can easily comprehend distinctions in program performance denoted by a limited number of distinct program scores.

We believe, however, that the composite score approach is problematic. Assigning weights to indicators involves an arbitrary decision about their relative importance and conveys a false sense of precision. Often the teacher impact measure is given the greatest weight because it is prima facie the most compelling evidence of the adequacy of professional training. It is arguable, however, that indicators assume varying degrees of relative importance depending upon the larger context. In a state with a severe shortage of teachers, for example, a program’s contribution toward easing the shortage may take on added importance. And there is a good deal of controversy over the importance of candidate academic strength as a program performance indicator because many teacher educators are concerned that an emphasis on academic competency at program admissions will reduce the already limited pool of minority teacher candidates (Coble et al., 2016). Moreover, as we noted previously, the teacher impact measure may not provide much distinction between the performance of programs on that indicator (Ehlert et al., 2013; von Hippel et al., 2016), and giving it added weight in a composite score would reduce scoring variation between the composite program scores as well. Also, methodological considerations expressed by researchers imply the need for caution in putting a lot of weight on teacher impact measures. Koedel and Parsons express the concern that teacher impact measures may not yet be reliable enough to be given more than moderate weight in a high-stakes accountability context (Koedel & Parsons, 2014). And Goldhaber (2013) suggests that there are added challenges to using teacher impact measures in the context of evaluating teacher preparation programs. Similar concerns would apply to other program performance measures, as well.

More likely than not, most programs will perform better on some indicators than others—with few programs that perform poorly or outstandingly on all of them. Such nuances are lost, however, when programs are assigned a single overall performance score. And programs have a legitimate reason to be resentful when their programs are tagged with composite scores that,
at bottom, are contrived arbitrarily or based on statutes or policies that contradict the cautions of expert researchers. In our work with states to develop these systems, we have seen several examples of the perceived over-aggressiveness of the system being met with a serious backlash from one or more stakeholder groups that not only slows progress but can fundamentally weaken the EPP accountability system.

Looking to the Future

There is much that we still don’t know about this relatively new enterprise of constructing annual, indicators-based accountability systems for assessing and improving the performance of educator preparation programs. We do not yet have sufficient empirical evidence to judge that the indicators commonly adopted in states’ efforts to develop these systems are the most suitable for identifying program strengths and weaknesses and for grounding a continuous improvement process. And any evidence we do have must take into consideration the specific measures employed to apply the indicators—measures that are frequently weak and vary from state to state. We do not yet know, in many cases, what scores on indicator measures—generally or in a specific state—reliably signal a serious deficiency in program performance or an exemplary strength. Is a cohort average score below the 50th percentile nationally on a Praxis II assessment, for example, a cause for concern that a program has inadequate content knowledge standards? Are the available tests of content knowledge of teaching sufficiently rigorous and focused on the right kind of knowledge even to assess adequate content knowledge for prospective teachers? Is a completer cohort two-year job placement rate of 60% an indication of a program weakness if the statewide average is 75%?

Some of these questions will eventually be answered by the individual states empirically. Indeed, research studies are underway in at least two states—Massachusetts and Tennessee—that may shed light on the validity of some program measures and the impact of accountability systems on program improvement. (Theobald, Conoway, Cowan, & Peske, in progress; Ronfeldt & Campbell, 2016). Questions about the right indicators and measures and about where to set the benchmarks for satisfactory performance will continue to be raised. And there will continue to be questions about how to shape accountability policies and practices so that they most effectively promote program improvement.

The measures available to gauge the performance of educator preparation programs will never tell the entire story—no matter how much better they become. The data conveyed by the measures constitute a signal, not a cause, and they must be supplemented by the knowledge and experience of teacher education experts and state officials—about their specific programs and about the field in general—to pinpoint the precise deficiencies of their programs and identify the steps that will be necessary to address them.
Creating a Data Culture  59

Every success in identifying those causes and strategies, however, can lead to an improvement of the signals, and thus in the diagnostic quality of the performance data that program accountability systems reveal.

At the end of the day the fundamental question is how well a given state’s accountability system is contributing to the increased effectiveness of preparation programs and of the novice teachers who graduate from them. A great deal of faith, effort, and money have gone into the development of these systems, but it will still be several years before even the most advanced state program accountability systems can demonstrate their ultimate value. Every one of the state education agencies we have worked with seeks to ensure that their program review and accountability systems go beyond the identification of programs that are low-performing and that they motivate and guide efforts of the providers themselves (with state assistance if possible) to identify the specific program deficiencies hindering performance and develop sound strategies to remedy them.

The data required for any truly effective program-level improvement process will have to go beyond the basic performance scores that the new program accountability systems report to the EPPs and (probably in more summary form) to other stakeholders. We believe, however, that a sufficiently comprehensive set of well-crafted performance measures based on valid and reliable data can help EPPs construct an extremely suggestive profile that can guide the search for root problems and solutions. The kinds of program performance measures that many states have adopted and that are reflected in the Key Effectiveness Indicators (KEI) are signals of possible program weaknesses or, for high-scoring programs, possible strengths. And we believe that an analysis of program scores on multiple measures within the several program performance domains identified in the KEI can provide both confirmation of program strengths and weaknesses and an initial indication of where program deficiencies might lie. If, for example, completers of one middle grades mathematics program have low value-added scores, relatively high scores on observed classroom performance, and low scores on the licensure content test in mathematics, it’s not too much of a stretch to hypothesize that their subject matter preparation may be weak.

A more accurate diagnosis, however, and ultimately an effective strategy for program improvement require data and contextual information that go well beyond what can be gleaned from state-required accountability measures. Although, for program accountability purposes, state officials may be interested, for example, only in overall completer impact on K–12 students’ learning, the programs themselves are likely to benefit from knowing whether completers’ students show patterns of strength or weakness on particular subsections of the achievement tests used to measure teacher impact. Programs’ self-knowledge, whether gained through a long-term state formal program review, national accreditation, a program inspection process, or some other self-study process, can provide valuable contextual and highly
specific information that can help programs home in on important strengths and shortcomings and develop effective improvement strategies.

An effective accountability system can play a vital role in both stimulating and guiding the program improvement process. A system with high standards will require programs to undertake continuous improvement and focus on the data that are likely to be most strategic for the improvement process. State officials can also give programs guidance for constructing effective continuous improvement plans and hold programs accountable for implementing those plans.

Indeed, already states are actively engaging their teacher preparation programs in a continuous improvement process linked to an annual program review and accountability system. And an increasing number of studies not only discuss the details of that sort of engagement but also provide analysis of the conditions that make it successful (Bastian, Lys, & Pan, 2018; Gansle et al., 2015).

We would hope that the process of developing strong educator preparation performance systems in the states will become a higher priority for the larger education community. Teacher educators need to be active and willing partners in providing guidance to the development of annual program reporting systems in their states. They need to accede to the demands for stricter accountability based on program outcomes and to recognize the importance of using accountability data in program improvement decisions. At the same time, they have an invaluable role in ensuring that the data and reports will be useful and in guiding state officials so that objective is met.

Researchers and psychometricians have the expertise required to refine the statistical models that generate performance measures, assist states in the development and implementation of both their data and accountability systems, and develop ever better assessments of teacher candidates’ relevant knowledge and skill and of teachers’ performance in the P-12 classroom. Working together, researchers and teacher educators should be able to mine the rich new data that will emerge from the program accountability systems and thereby increase the field’s knowledge of what policies and practices work in teacher preparation—and for which candidates and under what circumstances.

What ultimately may be possible, given a concerted and coordinated effort from the various contributors in the field, is using the most valid and reliable measures with high-quality data as a solid platform for refining the science and art of program improvement. The promise is more powerful and accurate discernment of the most serious problems hindering teacher preparation and effective engagement with teacher educators and other key stakeholders in both post-secondary and K–12 education. Better data and better policies and tools that make use of those data should enhance the recruitment, preparation, and support of teachers so that preparation programs truly meet the needs of districts and states for the teachers our children deserve.
Notes

1. The seven questions are as follows:
   1. What is the primary purpose of the teacher preparation program evaluation system?
   2. Which aspects of teacher preparation matter most?
   3. What sources of evidence will provide the most accurate and useful information about the aspects of teacher preparation that are of primary interest?
   4. How will the measures be analyzed and combined to make a judgment about program quality?
   5. What are the intended and potentially unintended consequences of the evaluation system for teacher preparation programs and education more broadly?
   6. How will transparency be achieved? What steps will be taken to help users understand how to interpret the results and use them appropriately?
   7. How will the evaluation system be monitored?

2. There is, of course, a robust literature on the strengths and limitations of value-added assessment of teachers, as well as the use of value-added measures to assess preparation programs. (See, for instance, the sample of related studies on the U.S. Government’s Science.gov website, www.science.gov/topicpages/t/tennessee+value-added+assessment.html#.) We believe that there is sufficient methodological and empirical support for states to consider the use of value-added assessment in a preparation program accountability system, but the success of such an effort requires valid and reliable assessments of student learning and the construction of solid value-added models.

3. It will not be equally easy to apply the measures to all programs, especially when it comes to tracking placement, retention, and performance of program completers. Some programs, for example, may send almost all their completers into private schools, charters, or out of state. And one variable that may require exceptions for some programs is program cohort size. Small programs may have to aggregate several years’ data from consecutive cohorts in order to obtain meaningful measures. In some cases, the performance of very small programs simply may have to be determined more qualitatively.

4. We address below the issue of a teacher candidate’s content knowledge being principally acquired in arts and sciences courses.

5. Wyoming, for example, has only one institution in-state (the University of Wyoming) that grants teacher licenses.

References


### Table 1 Key Effectiveness Indicators

<table>
<thead>
<tr>
<th>Assessment Categories</th>
<th>Key Indicators</th>
<th>Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>I Candidate Selection and Completion</td>
<td>Academic Strength</td>
<td>PRIOR PROFICIENCY—1. Average GPA of candidates in most recent coursework (high school or college) prior to program entry—TPP. 2. Overall entering cohort average percentile score in national distribution on standardized entrance tests required by IHE or EPP (SAT, ACT, GRE, MAT, or College Skills Test (e.g., Praxis Core)—TPP, EPP. COMPLETER PROFICIENCY—Completer GPA in required subject major courses compared to all university students in same major—TPP.</td>
</tr>
<tr>
<td></td>
<td>Teaching Promise</td>
<td>ATTITUDES, VALUES, AND BEHAVIORS SCREEN—Percent of accepted program candidates whose score on a rigorous and validated &quot;fitness for teaching&quot; assessment demonstrates a strong promise for teaching—TPP.</td>
</tr>
<tr>
<td></td>
<td>Candidate/Completer Diversity</td>
<td>DISAGGREGATED COMPLETIONS COMPARED TO ADMISSIONS—Number &amp; percent of completers in newest graduating cohort AND number and percent of candidates originally admitted in that same cohort: overall and by race/ethnicity, age, gender, etc.—TPP, EPP.</td>
</tr>
<tr>
<td>II Knowledge and Skills for Teaching</td>
<td>Mastery of Teaching Subjects</td>
<td>CONTENT KNOWLEDGE TEST—Program completer mean score*, tercile distribution, and pass rate on rigorous and validated nationally normed assessment of college-level content knowledge used for initial licensure—TPP. *Validated proficiency benchmarks may be substituted for mean scores on these assessments</td>
</tr>
<tr>
<td></td>
<td>Subject-Specific Pedagogical Knowledge</td>
<td>PEDAGOGICAL CONTENT KNOWLEDGE TEST—Completer mean score*, tercile distribution, and pass rate on rigorous and validated nationally normed licensure assessment of comprehensive pedagogical content knowledge—TPP. *Validated proficiency benchmarks may be substituted for mean scores on these assessments</td>
</tr>
<tr>
<td></td>
<td>Completer Teaching Skill</td>
<td>TEACHING SKILL PERFORMANCE TEST—Program completer mean score*, tercile distribution, and pass rate on rigorous and validated nationally normed licensure assessment of demonstrated teaching skill—TPP. *Validated proficiency benchmarks may be substituted for mean scores on these assessments</td>
</tr>
<tr>
<td></td>
<td>Completer Rating of Program</td>
<td>COMPLETER PERCEPTIONS OF PROGRAM QUALITY—State or nationally developed program completer survey of program quality and teaching preparedness, by cohort, upon program completion and at end of first year of full-time teaching—TPP.</td>
</tr>
<tr>
<td>III Performance as Classroom Teachers</td>
<td>Impact on K–12 Student Learning</td>
<td>TEACHER CONTRIBUTION TO STUDENT LEARNING—Success of program completers in 2nd and 3rd most recent cohorts or of alternate route candidates during their first two years of full-time teaching based on valid and rigorous student learning measures, including value-added or other statewide comparative evidence of K-12 student growth overall and in high-need schools. Average student growth score for completer cohort and percentage of completers in cohort scoring below the 33rd and above the 67th percentile compared to the average score and distribution for all novice teachers statewide and for all teachers statewide—TPP.</td>
</tr>
<tr>
<td>Demonstrated Teaching Skill</td>
<td>ASSESSMENTS OF TEACHING SKILL—Annual assessment based on observations of program completers’ or alternate route candidates’ first two years of full-time classroom teaching, using valid, reliable, and rigorous statewide instruments and protocols—TPP.</td>
<td></td>
</tr>
<tr>
<td>K–12 Student Perceptions</td>
<td>STUDENT SURVEYS ON TEACHING PRACTICE—K–12 student surveys about effectiveness of completers’ or alternate route candidates’ teaching practice during first two years of full-time teaching, using valid and reliable statewide instruments—TPP</td>
<td></td>
</tr>
<tr>
<td>IV Contribution to State Workforce Needs</td>
<td>Entry and Persistence in Teaching</td>
<td>EMPLOYMENT—Percent of completers from 2nd and 3rd most recent completer cohort (including alternate route completers) employed within two years of program completion, by gender and race-ethnicity—TPP, EPP. PERSISTENCE—Percent completers (traditional and alternate route) from the fourth most recent completer cohort who remain in teaching or other educational roles for one, two, and three years after initial entry. OR, percentage of completers attaining a second stage teaching license in states with multitered licensure—TPP, EPP. Percentages for Employment and Persistence for each program to be compared to statewide mean average for each certification field and mean average for programs in all fields. EPP average to be compared to mean average for all EPPs statewide.</td>
</tr>
<tr>
<td>Placement/Persistence in High-Need Subjects/Schools</td>
<td>EMPLOYMENT—Percent of completers from 2nd and 3rd most recent completer cohort (including alternate route completers) employed within two years of program completion in high-needs schools and subjects, by gender and race-ethnicity—TPP, EPP. PERSISTENCE—Percent completers (traditional and alternate route) from the fourth most recent completer cohort who remain teaching in high-need subjects or in teaching or other educational roles in high-need schools for one, two, and three years after initial entry—TPP, EPP. Percentages for Employment and Persistence for each program to be compared to statewide mean average for each certification field and mean average for programs in all fields. EPP average to be compared to mean average for all EPPs statewide.</td>
<td></td>
</tr>
</tbody>
</table>
Some of the people driving us all hard into the future on the back of new technologies appear to assume that if we all focus hard enough on information, then we will get where we want to go most directly. This central focus inevitably pushes aside all the fuzzy stuff that lies around the edges—context, background, history, common knowledge, social resources. But this stuff around the edges is not as irrelevant as it may seem. It provides valuable balance and perspective. It holds alternatives, offers breadth of vision, and indicates choices. It helps clarify purpose and support meaning. Indeed, ultimately it is only with the help of what lies beyond it that any sense can be made of the information that absorbs so much attention.

—Brown & Duguid, 2002, *The Social Life of Information* (pp. 1–2)

This chapter is about the “fuzzy stuff” that matters to making data useful and used for the improvement of teacher education. We begin by examining the contemporary policy contexts in which theories of action related to data and program improvement have become so prominent over the last decade. We then review thematic findings from research on data use from some related fields of human service and education that may be useful for understanding the challenges we are experiencing in teacher education. While research on data use practices in teacher education is relatively recent and limited in scope, we examine the existing studies and interpret them in terms of their implications for program-level data use practice. Finally, we offer recommendations for policy makers related to some of the problems of practice that have emerged in the context of the current intensification of accountability policy in teacher education.

Accountability Policy in Teacher Education: Mandates, Theories of Action . . . and Missing Agendas

Contemporary policy in teacher education is clearly accelerating its attention to issues of measurement, information systems for archiving and analyzing data, and data use (Council of Chief State School Officers [CCSSO],
Two distinct goals are apparent in these policies. The first focuses on program accountability, often with the explicit goal of evaluating and ranking (and potentially sanctioning) programs based on one or more measures of program quality and effectiveness (Crowe, 2010). The second goal focuses on using data for program improvement. Very often these goals are intertwined. For example, in a recent report on their effort to support development and use of improved data systems in teacher education, the Council of Chief State School Officers’ Network for Transforming Educator Preparation (2018) described its goals as:

Developing and building outcomes-oriented data systems for their educator preparation programs (EPPs) designed to: report program performance back to EPPs to promote continuous improvement, publicly report on EPPs to increase transparency of EPP performance, and review EPP performance for accountability purposes (e.g., to make program approval or renewal decisions).

(p. 3)

This theory of action, with accompanying standards for using data for program improvement, has also become a salient requirement for program accreditation in state departments of education and professional associations. For example, the Council for Accreditation of Educator Preparation (CAEP, 2013), the major professional accreditation organization for teacher education, requires in Program Standard 5.3 that:

The provider regularly and systematically assesses performance against its goals and relevant standards, tracks results over time, tests innovations and the effects of selection criteria on subsequent progress and completion, and uses results to improve program elements and processes.

(p. 14)

The salience of issues related to data and data use in state and national policy conversations has also led to a renewal of interest in problems of measurement and program evaluation in teacher education (Feuer, Floden, Chudowsky, & Ahn, 2013). An enormous effort has been undertaken to identify or develop valid and reliable measures for evaluating teacher education programs (e.g., Allen & Coble, this volume, Chapter 3; Noell, Brownell, Buzick, & Jones, 2014), and numerous state legislatures have projects underway to incorporate a variety of outcome measures into “data dashboards” designed to support the uses of data for both accountability and program improvement purposes (CCSSO, 2018).

Reliable and valid measurement of teacher education program outcomes is clearly an essential cornerstone of any theory of action related to data use. And the attention of policy makers has been consumed largely by these and
related challenges, gathered often under the general problem of developing “data systems.” While substantial technical progress is being made on this front, significant challenges remain. These involve underlying value conflicts regarding the goals of teacher education (Zeichner, 2014), practical difficulties in measuring teacher quality (Kane et al., 2013), and challenges in drawing causal inferences from correlational data evaluating teacher education program outcomes (Feuer et al., 2013).

In this chapter, we argue that an equally significant set of challenges for making data useful and used for program improvement lies in the domain of what Brown and Duguid (2002) have termed “the social life of information.” In their original identification of this problem, Brown and Duguid (both prominent cognitive scientists) were describing their own experiences with efforts underway at the Xerox Corporation to develop technical specifications for managing problems with machines—certainly a domain of activity where one might imagine that technical knowledge was sufficient to the tasks at hand. That turned out not to be the case. Unfortunately, in most current policy discourse in both K–12 and teacher education, the challenges of data use have been similarly represented as essentially technical issues, with accompanying inattention to the social and organizational dimensions of the work required to make emerging technical systems both useful and used (Jennings, 2012).

In this chapter, we attempt to construct a more holistic view of data use work, drawing on the framework proposed by McDiarmid and Peck (2007), in which data use practice is conceptualized to involve an interplay of the values, beliefs, and motivations of the people undertaking the work, the features and affordances of available data tools, and the features of the organizational contexts in which the work takes place. The theory of action underlying our use of this framework underscores the role of learning in the policy implementation process. Spillane, Reiser, and Reimer (2002) have articulated this view well: “From a cognitive perspective, a key dimension of the implementation process is whether, and in what ways, implementing agents come to understand their practice, potentially changing their beliefs and attitudes in the process” (p. 387). While our conceptualization of the learning process itself draws on a more sociocultural perspective, in which learning processes are mediated by conceptual and materials tools as well as the affordances and constraints of social interactions in the workplace (Billett, 2006; Engeström, 1987), we follow Spillane and colleagues (2002) in appreciating the power and importance of practitioner learning as a critical source of motivation and engagement with the work of systemic change and program improvement. With this in mind, we have used the People-Tools-Organizations (PTO) framework as a conceptual tool to support practitioner analysis and action related to their data use practices (Peck,
McDonald, & Davis, 2015). We expand our definition of the people, tools, and organizations dimensions of data use practice below.

**People**

Information systems in teacher education do not stand outside of the webs of meaning that program members construct about their work, both individually and collectively. We conceptualize this aspect of data use work to be about the people involved, including particularly the values, beliefs and aspirations that motivate their engagement (or disengagement) with opportunities to learn that are afforded by program outcome data.

**Tools**

This is the dimension of data use work that has received the most attention from policy makers, funders, and digital entrepreneurs. Clearly, the quality and accessibility of the various data tools, including measurement instruments, data warehouses, and “dashboards” for monitoring and analysis of program processes and outcomes, matter tremendously. Our interests in tools extend also to questions about accessibility and flexibility, particularly with respect to their use by faculty, program staff, and even candidates themselves. At issue here is the extent to which data tools become internalized, valued, and used within the local culture of a program as tools for program improvement, or whether they remain essentially instruments used on the program by others for accountability purposes.

**Organizational Policy and Practice**

A third dimension of the social ecology in which information systems and decision making in teacher education exist has to do with the “ways of doing things” that are negotiated in programs as organizations. Here we refer to program policies (e.g., compensation, tenure and promotion, job descriptions, etc.) and practices (meetings, data use gatherings, work routines, norms for collegiality) that affect participation in collective activities related to data use.

Figure 4.1 represents the interrelated nature of these dimensions of data use, including the disconnections between contemporary policy discourse focused primarily on tools and the other dimensions of data use practice.

Research on data use from both within and outside the field of education suggests the disconnection represented here is by no means unique to teacher education. Moreover, the “fuzzy stuff” appears to be highly consequential across arenas of social practice in which policy makers and practitioners attempt to use data for program improvement.
Social science research from multiple fields of human service, including nursing, rehabilitation, and social work, converges in demonstrating that data use involves much more than simply collecting and analyzing evidence on program (or treatment) outcomes (Estabrooks, Midodzi, Cummings, & Willin, 2007; Lipshutz & Popper, 2000; Sudsawad, 2007; Weiss, Murphy-Graham, Petrosino, & Gandhi, 2008). Particularly at issue are the ways in which organizational policies and practices shape opportunities for deliberation, learning, and action (Nicolini, Gherardi, & Yanow, 2003). Spillane and Miele (2007) have been particularly articulate in expressing this viewpoint in the context of data use in education:

For those in the trenches, our take home message is this: work practice is where the rubber meets the road in the schoolhouse. Understanding how information becomes evidence, and how this evidence gets used or goes unused requires attention to work practice. Work practice can be difficult to assess, and even more difficult to analyze. We often gloss over these difficulties by focusing on simplified strategies for evidence-based decision making. But to understand evidence use, we must attend to practice, which necessitates attention to interactions between people,
as well as to how these interactions are mediated by aspects of the situation (such as organizational routines and tools).

(p. 68)

Several findings from research on data use in the P-12 sector support this view. For example, a line of research conducted by Coburn and her colleagues (Coburn, 2001; Coburn, Toure, & Yamashita, 2009; Coburn & Turner, 2011) shows that data use is a highly interpretive process influenced by individual beliefs and knowledge. For example, people often attend primarily to data that support their personal beliefs and ignore or discount evidence that might contradict or challenge their beliefs (Coburn & Talbert, 2006; Ingram, Louis, & Schroeder, 2004). Moreover, people with different beliefs may interpret the same evidence in contradictory ways (Coburn, 2001; Coburn et al., 2009).

This general finding underscores the importance of a second theme from this literature regarding the functions of collaboration in data use activities. Based on their study of data use in district central offices, Coburn and colleagues (2009) argued that substantive “cross-divisional engagement is likely necessary to develop sets of understandings about instructional issues that are truly shared and that, in turn, can facilitate joint definitions of problems and shared visions of solutions” (p. 1146). These studies and others (Hamilton et al., 2009; Nelson & Slavit, 2007; Wayman, 2005; Young, 2006) suggest that distributing opportunities, responsibilities, and supports for engagement among a variety of stakeholders, including teachers in the P-12 context, is important to the success of interventions promoting data use.

Developing and sustaining collaborative data use practices also requires strategic organizational support (Anderson, Leithwood, & Strauss, 2010; Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006). For example, Datnow, Park, and Wohlstetter (2007) found that “high data use” schools and districts built capacity for data use at the school level by creating structured time and opportunities for collaboration both within and across schools. Other studies have shown the importance of providing adequate time to examine data or conversely identified lack of time as a barrier to data use (e.g., Coburn, 2001; Ingram et al., 2004; Means, Padilla, & Gallagher, 2010; Wayman & Stringfield, 2006).

Taken together, studies from other fields of human service, as well as those from education, clearly suggest that technical advances in data systems must be accompanied by strategic social and organizational supports if teacher educators are to be reasonably expected to take up opportunities to use data to improve their individual and collective practice. Fortunately, recent research in teacher education is beginning to offer some guidance about how teacher education practitioners can be supported in engaging new opportunities to use data for meaningful program improvement. These studies suggest that many of the strategies and practices developed in the P-12 sector have considerable value in the contexts of teacher education (Datnow et al., 2007; Kerr et al., 2006). In other cases, the unique history and context of teacher education programs, particularly those situated in institutions of
higher education (Fraser, 2007; Goodlad, Soder, & Sirotnick, 1990), suggest that alternative strategies may be important for motivating and supporting faculty and staff engagement with data use activities (Davis, 2016, 2018).

Research on Data Use in Teacher Education

We begin by noting that the assumption that data use in teacher education matters itself relies on a complex set of inferences and assumptions that are certainly rational but that have limited empirical support at this point in time. This state of affairs appears unlikely to change soon, when considering the methodological challenges involved identifying data use practices that lead to programmatic improvements that in turn affect the practice of teacher candidates as they enter the classroom and that are then maintained in their practice as novice teachers with sufficient fidelity as to improve educational outcomes for the children in their classrooms.

Several studies do support the assumption that data use can matter to efforts at program improvement in teacher education. For example, in one early and comprehensive report, Reusser, Butler, Symonds, Vetter, and Wall (2007) describe how multiple data sources, including both standardized tests of subject matter knowledge and teacher performance assessments, were used to identify program weakness and guide efforts at improvement in their program. Using a Total Quality Management model, faculty and academic leaders collaborated to identify needed changes in advising, coursework content, and practicum policies. Follow-up measures indicated these changes were associated with continuous improvement in both subject matter knowledge test pass rates and teaching performance assessment scores over a five-year period. The authors interpreted observed program changes to be related not simply to data use per se but to a broader change in organizational culture within their program.

To our knowledge, the only study to date which has even attempted to span the entire chain of causal inferences related to assumptions underlying contemporary data use policies is a study recently reported by Gansle and her colleagues (Gansle et al., 2015). In this study, standardized tests of academic achievement by students enrolled in the classrooms of teacher education program graduates were used in a statewide assessment of the “value-added” by specific preparation programs to the teaching effectiveness of their graduates. These measures indicated that students in the classrooms of graduates from one program were not making expected progress on standardized tests of reading achievement. Administrators in this program subsequently undertook a vigorous effort to improve coursework related to teaching reading, including adding new literacy coursework, hiring a more highly qualified literacy instructor for the program, and implementing direct observational measures of candidate teaching practice that were used as interim measures of instructional improvement. Subsequent value-added assessments showed significant improvement in student literacy achievement over a two-year period. While this report offers some valuable
Building Capacity and Commitment

support for the theory of action underlying most contemporary policies related to data use and program improvement, it is clearly subject to all of the threats to internal validity associated with nonexperimental case studies.

With the limitations of these studies and others (e.g., Cochran-Smith & the Boston College Evidence Team, 2009; Peck, Gallucci, & Sloan, 2010) in mind, we argue that the existing research suggests that data can provide valuable motivation and guidance for program improvement in teacher education. In the next section, we use the People-Tools-Organizations framework (see Figure 4.1) to review existing research in an attempt to summarize what we currently know about the conditions under which this takes place.

What Kinds of Tools Are Useful?

While we have argued here that tools are not the only aspect of data use practice that matters, that assertion should by no means be interpreted to suggest that they are not important. Data tools, including the types of data collected, the ways they are organized and archived for analysis and retrieval, and the ways they are presented, seriously affect what program outcomes can be “seen” and subjected to critical analysis and deliberation as objects of potential change. Existing studies underscore several important issues about the affordances and constraints of tools related to data use.

First, a point of convergence across several comprehensive studies of data use is that using multiple measures is important to triangulate perspectives and findings about program outcomes (Cochran-Smith & the Boston College Evidence Team, 2009; Davis & Peck, 2018). Two studies reported by our own group underscore this point (Peck et al., 2010; Peck & McDonald, 2014). In these investigations, we observed that the introduction of new data sources provided new perspectives on program outcomes that had been obscured by prior measures, including highly positive results from graduate satisfaction surveys and high national rankings for the programs involved. These findings based on new data sources were pivotal in disrupting faculty assumptions about what their students were learning and in motivating faculty engagement with program improvement efforts.

While the notion of using multiple sources of data is widely accepted, and even mandated in many state policies and accreditation standards (cf. CAEP Standard 5.1), this does not imply that all types of data are equally useful, nor that they serve the same functions in program improvement work. In our most recent and comprehensive study, our research group investigated this question (and a variety of additional issues around data use practices) in 10 “high data use” programs situated in diverse state policy contexts and with various institutional missions and sizes across the country (Davis & Peck, 2018; Peck et al., 2015; St. John, Davis, & Peck, 2018). We asked academic leaders, faculty, and P-12 partners in these programs about the sources of data they found useful for their program improvement work. We found that data from teaching performance assessments (e.g., the edTPA [Stanford Center for Assessment, Learning and Equity, 2013], the Teacher
Work Sample [Schalock, 1998], and others) were consistently viewed by practitioners to be the most useful data sources for identifying specific candidate performance outcomes of concern and in providing guidance for program improvement efforts.

Findings from several studies suggest that the value of TPAs does not lie solely in the analysis of composite scores or even results on individual rubrics but rather in qualitative analysis of teaching artifacts (e.g., video, lesson plans, samples of P-12 student work) collected through the assessment process (Bunch, Aguirre, & Tellez, 2009; Devlin-Scherer, Burroughs, Daly, & McCartan, 2007; Peck, Gallucci, Sloan, & Lippincott, 2009; Sloan, 2013). For example, Bunch and colleagues (2009) used artifacts from the Performance Assessment for California Teachers (Pecheone & Chung, 2006) to investigate what could be learned about the ways candidates in their program articulated and enacted their understanding of the needs of English learners in their classrooms. A specific issue of interest to Bunch and colleagues was the function of the portfolio artifacts themselves (which included lesson plans, videos of classroom instruction, and samples of P-12 student work), rather than rubric scores, as sources of information about candidate learning. Based on detailed analysis of portfolio artifacts from eight teacher candidates with high proportions of Hispanic students in their classrooms, the authors found that the portfolio artifacts from the teaching performance assessment allowed them to identify both strengths and limitations in candidate’s interpretations and responses to their English learner students’ instructional needs. The authors identify a variety of ways in which these data, and other analyses of the performance assessments, could be used to enrich teacher educators’ understanding of what candidates were learning in the program and, ultimately, to improvement of teacher preparation.

In a related study, Devlin-Scherer and colleagues (2007) used data from the Teacher Work Sample (Schalock, 1998) to identify areas of candidate teaching strength and weakness, with related implications for program improvement. Using a detailed analysis of teaching performance artifacts from eight candidates in their secondary program, Devlin-Sherer and her colleagues identified priorities for program improvement, including the need to better define how community context can be helpful, strengthen connections between coursework and classroom applications, and expand the repertoire of classroom assessment techniques candidates employed in their evaluations of student learning. These authors also described ways in which the performance assessment process itself offered a potent context for candidate learning and improvement of practice.

Qualitative analysis of TPA artifacts is clearly a time-intensive process, which raises issues about how to select relevant cases from large archives of candidate work for follow-up analysis. In some contexts, such as that reported by Bunch et al. (2009), case selections may be driven by specific programmatic questions that imply focus on specific cases. However, Bastian, Lys, and Pan (2018) have recently developed an empirical approach to
case selection, using latent class analysis to identify patterns of TPA rubric scores that suggest specific issues and related cases for deeper, qualitative analysis. These analytic methods appear promising for identifying priorities for program improvement, including more detailed analysis of specific case examples, that become evident from data collected and analyzed from large groups or cohorts of candidates.

In the Davis and Peck (2018) study, we also investigated our informants’ views about the value of other data sources, including value-added measures of program outcomes (Plecki, Elfers, & Nakamura, 2012). While practitioners interviewed for this study were generally skeptical of value-added data, there was also evidence in some programs that value-added measures had served an important “fire alarm” function—drawing program members’ attention to performance issues with their graduates that had not been previously identified (Gansle et al., 2015).

Practitioners in these programs (Davis & Peck, 2018) generally reported finding relatively little value (for program improvement purposes) in the kinds of graduate and/or employer satisfaction surveys that are widely used in the field. Their observation was that items on these measures often were so global as to make responses difficult to attribute to specific program features and practices. An additional concern reported in some programs was that the generally positive reputation of the program operated as somewhat of a “halo” effect, generating positive ratings on most survey items. In some cases, positive satisfaction ratings reportedly led faculty to overlook substantive problems with specific program outcomes, often leading to surprise when these problems surfaced on other measures more intimately linked with classroom practice. An example of this was also reported in the program described in Peck and colleagues (2010). This program was rated very highly by both graduates and employers with respect to its efforts to prepare teacher candidates to understand and engage the needs of English language learners (ELLs). However, upon examination of TPA measures of actual classroom practice, program faculty identified several issues of concern, particularly with respect to the extent to which candidates were able to integrate their knowledge of ELL issues into their instructional practices in content areas such as literacy and math. These new data led faculty to undertake a series of actions related to improving the integration of coursework in math and literacy methods with practices formerly taught primarily in coursework related to ELL concerns (Hill-Bonnet, Lippincott, & Scalzo, 2010).

Information technologies for collecting, archiving, analyzing, and reporting program data are important tools for making data useful and used (Peck & McDonald, 2014; Peck et al., 2015). Many, if not most, programs around the country are making considerable investments in the development of digital data platforms designed to facilitate collecting, aggregating, and analyzing program input, process, and outcome variables. In our study of high data use programs (Davis & Peck, 2018; Peck et al., 2015; St. John et al., 2018), we observed that leaders in these programs paid careful
attention to involving faculty users in designing new data systems, with a concomitant investment in professional and organizational development related to effective use of the new data systems to support activities related to program improvement (Cho & Wayman, 2014). In one program we observed that considerations around accessibility for some data analysis functions were extended to teacher candidates themselves, who were actively supported in using their individual data for monitoring, evaluating, and supporting their own progress in the program. An important observation related to the practice of making program data platforms broadly accessible (rather than making these tools accessible only by program assessment staff or institutional researchers) was that this practice facilitated broader and more active engagement of program faculty and staff and made data related to needs for program improvement more visible to the people who were ultimately responsible for implementing program change.

**Organizational Supports for Data Use Work**

One of the robust findings across data use literatures from multiple fields is that implementing and sustaining data use practices related to program improvement requires developing strategic organizational supports for changing work practice (Crosson & Bedrow, 2003; Marsh, 2012; Weiss et al., 2008). This has been clearly evident in our own research on “high data use” programs (Davis & Peck, 2018; Peck & McDonald, 2013, 2014). Several organizational policies and practices appear important to supporting this work.

First, several of our own studies (Davis, 2016, 2018; Davis & Peck, 2018; Peck et al., 2015; Peck et al., 2010; Peck & McDonald, 2013, 2014) as well as findings from other comprehensive studies of data and program improvement (Cochran-Smith & the Boston College Evidence Team, 2009; Reusser et al., 2007) suggest the importance of creating a highly inclusive approach to data analysis and program improvement deliberation and decision making. This appears to us to be particularly important to change initiatives in higher education, with its notoriously hierarchical culture (Goodlad, 1993). The principle here is that the people who are expected to implement changes to improve a program must be involved in interpreting data related to the need for change and designing feasible actions related to goals for improvement. On a practical level, this means developing policies that support the participation of program members, such as field supervisors and cooperating teachers, who typically provide crucial supports for candidate development but who are very often marginalized in program decision making about curriculum and instruction. For example, in some of the programs we have studied, “faculty” meetings were historically scheduled at times during the regular school day—times in which field supervisors were typically expected to be out in schools observing and coaching teacher candidates. Creating shared time and space for field-based and campus-based teacher
educators (including cooperating teachers) to jointly analyze and deliberate the implications of program outcomes involved rescheduling these meetings. In many cases this also necessitated making concomitant changes both to formal policies (e.g., job descriptions, compensation policies) and to informal norms and expectations (e.g., particularly those related to valuation of field-based perspectives on the work of the program).

Other organizational supports and incentives for data use work include changes to regular work processes such as faculty meetings, faculty performance reviews, and decision processes for initiating new programs and practices. For example, one of the “high data use” programs we visited in the Davis and Peck (2018) study had adopted an evidence-based approach to decision making regarding initiation of major changes in program structures, instructional practices, or curriculum, which required collection of pilot data demonstrating the potential value of proposed changes before these were adopted program-wide. In all of the programs in this study, we were unsurprised to find that making “time” was viewed as a particularly serious and ongoing challenge to the development of new data use practices. We observed that the programs that were most successful at engaging this challenge became adept at embedding some aspects of their data use work into ongoing work practices such as faculty meetings, program reviews, and external reporting requirements. For example, one program we studied developed a yearlong calendar in which they planned specific data use activities that would take place during their regularly scheduled faculty meetings. Another program adopted the practice of suspending both regular coursework and fieldwork supervision for one week each spring in order to allow both field supervisors and course faculty to participate in collaborative scoring of performance assessments (Sloan, 2013). As might be expected, the initiation of this practice generated considerable concern from both course instructors and supervisors. However, over time, program members came to view the results of this reallocation of time to collaborative data analysis to be well worth the trade-offs.

In some cases, emerging opportunities and needs related to data use require the creation of entirely new work processes, with accompanying policies and related organizational supports for collaborative data-grounded work. One example of this we observed in several of the “high data use” programs described in Davis and Peck (2018) was the development of daylong data analysis and planning workshops variously referred to as “data summits,” “data days,” and “data retreats.” In these events, which typically occurred on an annual or biannual schedule, faculty, field supervisors, school partners, and academic leaders examined larger bodies of program data, including results of teaching performance assessments, graduate and supervisor satisfaction surveys, and (where available) value-added measures of student achievement in the classrooms of program graduates. We observed that academic leaders invested considerable planning time in these events, anticipating the need to reduce data sets to manageable size and to
provide thoughtful scaffolding for data analysis and action planning decisions. Data days offered valuable opportunities for more in-depth data use work across program members, sometimes including P-12 colleagues. Our observation was that careful planning and orchestration of these events was crucial to their success and that the experiences participants had at these events played a significant role in their views of the value of their individual and collective investment of their time in data use activities.

People: Motivating Faculty Engagement with Data Use Practice

Each of our own studies (Davis, 2016, 2018; Davis & Peck, 2018; Peck et al., 2010; Peck & McDonald, 2013, 2014; St. John et al., 2018), as well as reports from those undertaken by others (e.g., Cochran-Smith & the Boston College Evidence Team, 2009; Sloan, 2013), suggests that faculty interpretations of the purposes of data use—influenced by program leaders’ visions, goals, and articulated values around data use—are pivotal to their engagement with the work. At the core of this issue is the distinction between inquiry and compliance orientations to the work (Peck et al., 2010). The idea of “inquiry” here is grounded in perceptions related to locus of control and “ownership” of practice, which in turn are linked to larger questions about professional and institutional identity. Compliance motivations, in contrast, are associated with pressures to “use data for program improvement” emanating from external authorities such as policy makers, funders, and accreditation bodies (see McDermott & Lave, 2006, for a richer and more historical consideration of these and related issues about motivation, work, and learning).

Perceptions and related tensions about the purposes of data use are also important to teachers in P-12 settings (Finnegan & Gross, 2007; Hargreaves & Braun, 2013), as they are to practitioners in fields as diverse as medicine and military service (Lipshutz & Popper, 2000). However, ideological commitments to intellectual independence, autonomy, and inquiry, often gathered under the heading of “academic freedom,” are clearly central to the identity of university faculty, including those involved in the work of teacher education (Pullin, 2004). Whatever one’s views about the extent to which these features of university culture are desirable, these are the conditions under which the meanings of data use must be negotiated if university faculty and staff are to be engaged in productive program improvement work. Findings from the emerging literature on data use in teacher education suggest some strategies for managing these tensions (e.g., Cochran-Smith & the Boston College Evidence Team, 2009; Davis, 2016, 2018; Peck et al., 2010; St. John et al., 2018; Sloan, 2013).

Several studies suggest the importance of aligning data use work with faculty values (Cochran-Smith & the Boston College Evidence Team, 2009; Davis, 2018; Peck et al., 2010). For example, Cochran-Smith and her colleagues (2009) describe a careful process through which they engaged their
Building Capacity and Commitment

colleagues in defining local program goals related to social justice teaching, followed by an extensive effort to develop and use several program outcome measures for program improvement work grounded in these values. They underscored the importance of framing their work around “an exploratory and local approach to evidence construction,” which they conceptualized as “standing in marked contrast to the confirmatory and predetermined approach often involved in teacher education accreditation reviews, where the goal is to verify compliance with external standards” (Cochran-Smith & the Boston College Evidence Team, 2009, p. 463). Careful and strategic attention to local faculty values was also reflected in the approach Peck et al. (2010) used to manage tensions associated with the state-mandated implementation of a new teaching performance assessment. In this study, program leaders worked with faculty to undertake a critical analysis of state policy texts, locating areas of convergence and dissonance with locally “valued outcomes” for their program. A list of these valued outcomes was then developed and used to align and prioritize data collection and analysis activities, with an ensuing effort to make specific program improvements aligned with articulated faculty values.

The importance of establishing an inquiry orientation for data use work was also reflected in our study of “high data use” programs (Davis, 2018; Davis & Peck, 2018; Peck et al., 2015; St. John et al., 2018). In these programs, leaders carefully and strategically reframed accountability in ways that focused on local goals, values, and agency related to inquiry and program improvement. In a follow up investigation of leadership practices observed in three of these programs, St. John, Davis, and Peck (2018) report comments from one dean they interviewed who was particularly articulate about this idea:

If you’re doing it out of fear, the outcome is not going to be positive . . . you’re not going to be doing the things that are going to move your institution forward because you’re going to be so worried about whether you’ve done what they’ve wanted you to do. . . . I’m not responding to pressure. I’m not doing it because Kate Walsh is breathing down our necks or the federal government is disappointed in us. I’m doing it because I believe it’s the right thing to do.

(p. 18)

St. John and colleagues reported further that this framing of local motivations underlying using data for program improvement was distributed across many faculty and field supervisors, as well as individuals in formal positions of leadership (Sloan, 2013).

In a related analysis of these “high data use” programs focusing on how programs broadened engagement in data use, Davis (2016, 2018) found that program leaders (and others) engaged in particular relational practices, termed bridging practices, that helped align and support individual and collective motives, knowledge, responsibility, and agency. These practices
involved articulating and demonstrating individual and collective benefits of data use, developing shared responsibility for program outcomes, allowing variation in individual engagement, and addressing resistance respectfully and productively. The financial, human, and structural supports these programs created made data use work more accessible to program members from a practical standpoint (Davis & Peck, 2018), but the work program leaders did to clarify the values and goals of collaborative data use activities made the relevance and importance of the collective action requested from program members visible and compelling and created an evolving sense of purpose around data use work (Davis, 2016, 2018).

We believe that the motivational dynamics associated with the inquiry/compliance issue represent the most robust and important finding across existing studies of data use in teacher education (Cochran-Smith & the Boston College Evidence Team, 2009; Davis, 2016, 2018; Davis & Peck, 2018; Peck et al., 2010; Peck & McDonald, 2013, 2014; Sloan, 2013). In contrast to stereotypes about “faculty resistance to change” that have characterized many contemporary commentaries about university-based teacher education, faculty and staff in these studies were observed to engage vigorously and effectively in efforts to use data to improve their programs. Our interpretation of this finding is that, while the meanings of “data use” in the context of contemporary accountability policies are certainly important in other sectors of education (Hargreaves & Braun, 2013; Jennings, 2012; Cho & Wayman, 2014, they are absolutely pivotal to the responses to these initiatives by faculty and academic leaders in higher education settings.

Orchestrating Change: Engaging Problems of Practice Related to Data Use

Clearly, the challenges of building and sustaining capacity and commitment to using data for program improvement do not present themselves in the neat categories of “people, tools, and organizations.” In this section, we describe some of the ways in which these dimensions of data use practice are integrated in addressing the problems of practice we have observed repeatedly in the programs we have studied. We think of these as “promising” rather than “evidence-based” practices.

Getting Started

A challenge that immediately confronts academic leaders is how to initiate change processes related to data use. We have observed the following kinds of strategies employed effectively in one or more of the programs we have studied or research reports we have reviewed:

- Start small. Initial efforts to use data for program improvement are often most productive when they are small enough to be managed as
“pilot studies” rather than large systemic change projects. It is often helpful to situate initial efforts in the context of an existing “community of practice”—that is, a small group of faculty and/or staff that have a history of collaborative work with one another. This “pilot” group can report recommendations back to the larger program based on what they have learned, which can then be the basis for planning more expansive efforts (Peck et al., 2010).

- Clarify and prioritize outcomes and related data sources that are highly valued by program members. It is wise to begin building both capacity and commitment to data use by focusing on data sources that are close to the ground-level work of program members. Artifacts from teaching performance measures are often a good place to start, as these data often provide clear and concrete evidence of what candidates are taking up from coursework and fieldwork experiences (Bunch et al., 2009).

- Develop clearly articulated “framing” language clarifying the purposes of data use work. The most important issue here is to focus on values related to local inquiry rather than compliance with mandates or incentives from funders, state policies, or accreditation standards. It may be useful to align the work with the need to engage external pressures, but the core motivation for meaningful data use work should focus on improving the integrity and effectiveness of both individual and collective practice.

Managing Dissent

Major change initiatives almost always generate some degree of conflict, counterargument, and pushback. There is a very real temptation to view these responses as “resistance” and to treat dissenting program members as “problems.” Here are some alternative perspectives we believe are more productive ways of engaging dissent:

- Concerns about change expressed by faculty and staff are almost always grounded in substantive practical dilemmas, value-tensions, and loss. Inevitably concerns of this nature will be surfaced, whether it be in public meetings or in the hallways after those meetings. Our observation has been that inviting full and authentic deliberation of major change initiatives is an important part of building strong commitment to ensuing program decisions. It is important that dissenting program members experience respectful attention and engagement with their views, particularly by administrators and other program leaders, if program deliberations are to lead to decisions that are implemented rather than subverted (Davis, 2016, 2018; Davis & Peck, 2018).

- It can be very useful to collect data on how faculty are perceiving and responding to change proposals. One relative easy way to do this is to ask program members to complete anonymous “free writes” during or
after large group meetings in which controversial issues are deliberated. These data can be a valuable source of information about both the general views of the group and the nature of outlying concerns. The data can then be thematically analyzed, presented back to the group, and used as a basis for planning next steps of the change process in a way that reflects both respect for the views of program members and a commitment to “moving on” once reasonable consensus has been achieved (see Peck et al., 2010).

**Making Time and Space**

We have yet to meet a teacher educator who did not feel their “plate” of professional responsibilities was full . . . and overflowing. For this reason, change initiatives that are “added on” to existing workloads are likely to be unsustainable over time. Finding ways of integrating data use work into ongoing organizational routines and practices is essential. We have observed some strategies that appear useful to this goal.

- Data use work should be proactively scheduled into regular program meetings across the program year with the same regularity as other program functions, such as admissions or annual faculty review. This avoids the stress and extra labor of having to schedule ad hoc meetings—a practice which often contributes to a sense of the work being an “add-on” to regular duties.
- Specific times may be identified for releasing faculty and staff from regular duties for specified events (such as co-scoring teaching performance assessment portfolios) in order to support data use work (Sloan, 2013).
- We have noticed that careful planning of data use events and activities is extremely important to the experiences program members have with these activities. Important planning tasks include identifying priority data issues of importance to program members, preparing relevant data sets in ways that make them readily interpretable, strategically designing deliberative processes related to potential actions, and identifying concrete plans for follow-up evaluation of actions taken. A mistake we have seen many groups make in planning these events is trying to present too much information—which can quickly overwhelm the capacity of a group to thoughtfully process the evidence and effectively use it as a resource for decision making.

**Sustaining Changes in Practice**

Systemic changes in practice inevitably present challenges related to sustainability, and data use is no exception (Hubers, Schildkamp, Poortman, & Pieters, 2017). Proactive design around this issue is important. Several practices appear promising.
• Developing program documents with specific policies to support the work. These may include changes in written policies such as job descriptions, meeting schedules, and representation in program leadership teams, as well as clearly articulated program values around inquiry and evidence-based decision making that are written into program documents such as mission statements, websites, and other publications.

• Distributing leadership roles and responsibilities. Programs that rely extensively on the leadership of one person, often a dean or department chair, are vulnerable when leadership changes—as it inevitably does. Sharing opportunities and responsibilities for leadership broadly across multiple levels of program faculty and staff, as well as P-12 school partners, will help embed values and practices around data use into the “culture” of the program in ways than make them more durable across changes in leadership (Sloan, 2013).

• Strategic transition planning. When major leadership changes do occur, proactive transition planning focused specifically on maintaining core program values and related practices can help support continuity across leadership regimes. The regular inclusion of P-12 school partners in data use work, including this kind of transition planning, can provide systemic strength and resilience when leadership changes occur.

Policy Dilemmas and Directions

We opened this chapter with the observation that teacher education policies often conflate goals for program accountability with those for program improvement. The problems associated with these kinds of policies are well known (Hargreaves & Braun, 2013; Solberg, Mosser, & MacDonald, 1997; Weiss, 2012). For example, over two decades ago Solberg and his colleagues (1997), commenting on their work related to improvement of health care practices, observed that:

We are increasingly realizing how critical measurement is for the improvement we seek, yet how counterproductive it can sometimes be to mix measurement for accountability with measurement for improvement. Considered one by one, measurement for each purpose can be good and very important. If done poorly, it can be bad. If the measurements are mixed together in inappropriate ways, they can indeed be harmful or destructive, with the mixed purposes interfering with one another.

(p. 135)

Our own studies of teacher educators’ efforts to use data for program improvement suggest these problems are quite real in our field (Davis & Peck, 2018; Peck et al., 2010; Peck et al., 2015). Contemporary accountability
policies in many states now feature public comparisons of program outcomes based on measures of questionable validity. At the same time, both state policies and national accreditation standards routinely require that these and other data be used for program improvement—thus offering pressure but neither guidance nor support in responding to the complex challenges of organizational change implicit in these mandates. Our argument here is not that policies aimed at increasing the use of data in decisions about program policy and practice are bad—we continue to believe that data are a crucially important tool, and perhaps our most important tool, for improving programs. Our point is that the current policy environment appears more likely to produce the kinds of strategic deflection of what are interpreted to be “compliance” mandates, as described by Kornfeld, Grady, Marker, and Ruddell (2007), than to induce productive engagement with opportunities for learning and program improvement. We believe a more productive approach to policy would emphasize the development of organizational policies and practices related to building capacity and commitment to inquiry-oriented program improvement. Such an approach would focus less on producing artifacts documenting that data are collected and acted upon and more on questions such as “what (and how) are we learning about how to use data for program improvement?” Such an approach seems particularly appropriate in view of the paucity of empirical evidence that high pressure accountability policies, including those mandating data use, do in fact produce improved program outcomes.

Research Needs

Although computer data systems can support changes of practice, we found that agency for change rested in people, not in the technologies themselves. Indeed, teachers’ sensemaking about “data” and “data use” shaped whether and how systems were used in practice.

—(Cho & Wayman, 2014, p. 1)

We return, finally, to questions about the “fuzzy stuff” and to our concerns about missing agendas in policy research related to data use in teacher education. Again, we underscore that our arguments are not intended to detract from the value and importance of the work underway related to development of new tools for collecting, archiving, and analyzing data in teacher education. Nor are we arguing that program comparisons are inherently problematic, much less that such comparisons cannot be used as a valuable resource for learning and improvement. Our argument is that the attention of both policy makers and researchers has largely overlooked major domains of inquiry and development (conceptualized here to concern “people” and “organizations”) that must be engaged if we are serious about making data
matter to decisions about practice. Several specific research priorities appear particularly important at this point:

- Longitudinal studies of program development would be useful to building knowledge about how data use practices are initiated, developed, and sustained over time.
- More detailed process studies (e.g., Hall & Horn, 2012) would help us understand how teacher educators negotiate interpretations of data and make decisions about change. These studies would be particularly useful to the extent they incorporate multiple stakeholders, including particularly P-12 school partners.
- Case studies that provide more detailed description of the full cycle of inquiry, action, and P-12 student impact that is implicit in data use policies (e.g., Gansle et al., 2015) would help us understand the conditions under which data actually do influence programmatic changes that affect candidate practice, which then affects P-12 student outcomes.
- Focused investigation is needed related to issues arising from the introduction of “data dashboards” as tools for supporting evidence-based decisions about program improvement. The rapid development and adoption of these tools appears to be accompanied at present by relatively little empirical attention to questions about what kinds of data are included in these and whether and how they are used.
- Research on leadership in teacher education would be helpful, particularly with respect to the kinds of change processes that are the implied outcomes of current policy (Clement, 2014). To our knowledge, there is to date only a single study in the field addressing leadership issues in teacher education (Sloan, 2013).
- Studies of efforts at using data to improve program practice at scale would be useful, particularly those exploring the potential of networked improvement communities and similar collaborations between institutions as mediums for interorganizational learning and change (Bryk, Gomez, Grunow, & LeMahieu, 2015).

**Coda**

In appraising the current state of research, policy, and practice, we conclude by affirming our faith that data can, and should, matter to decisions we make in teacher education. They should matter more than they currently do. But we choose the word faith carefully here. Much of the work we have reviewed in the present chapter reflects good faith efforts by teacher educators to take up the challenges of making data useful and used for improving their programs. But we also notice that these efforts are not yet sustained by evidence of their effects so much as by faith in their possibilities. And we note with real concern that the attention and investment of policy makers,
researchers, and funders appears to reflect a remarkable disengagement with the people who do the work of teacher education and with the features of organizational policy and practice that shape that work. We have faith that, over time, the field will develop a richer and more complete understanding of how data can be used to improve practice, and that we will collectively improve our capacity and commitment to learning from experience.

Note

1. In this section we draw very directly on the knowledge and experience of our colleagues Kristen Cuthrell, Pat Lubke, Diana Lys, Desiree Pointer-Mace, Jen Scalzo, and Tine Sloan, with whom we collaborated in developing a set of practice-oriented resources for leadership and program development related to data use (see Peck et al., 2015).

References


Moving Toward Common Measures to Accelerate Improvement
A Roadmap for Educator Preparation Programs

Tracey Weinstein and Charis Anderson

Introduction
Over the last decade, shifts in state and federal policy and the introduction of new national accreditation standards have put increased pressure on educator preparation programs (EPPs) to produce more credible evidence of program quality (Council for the Accreditation of Educator Preparation [CAEP], 2013a; Title II of The Higher Education Opportunity Act, 2008). These shifts in the policy and accountability landscape have required EPPs to invest significant financial and human capital resources in collecting vast amounts of data on program inputs and outcomes (U.S. Government Accountability Office, 2015; Wilson, 2014). New accreditation standards have also required EPPs to find ways to leverage summative data collected for accountability for the formative purpose of improvement (CAEP, 2013b).

The theory of action underlying current accountability policies assumes, in part, that with access to more credible data, EPPs will improve the quality of preparation provided. This theory of action rests on several important assumptions, including: that data collected for accountability are useful for improvement; that EPPs have the necessary capacity (e.g., organizational structures and supports including financial and human capital resources) to use these data to drive improvement; and that EPPs know what to change within their programs to improve in meaningful ways. Our work with EPPs across the country, however, suggests that these assumptions may not always hold, making it unlikely that accountability alone will encourage the improvements in program quality necessary to ensure future teachers are consistently prepared to meet the needs of all students.

First, an inherent tension exists between data collected for accountability and data for improvement (McDiarmid, forthcoming). To meet accountability requirements, EPPs must demonstrate how they are meeting standards, a posture that naturally encourages them to collect and put forward evidence that positions their programs in the best possible light (and to minimize data that does not). A true improvement effort, however, requires just the opposite: to improve, EPPs must adopt an inquiry stance toward evidence that lays bare both strengths and weaknesses. Further complicating matters is
that data collected for accountability are often at a grain size that isn’t useful for improvement (Data Quality Campaign, 2017). For example, EPPs often receive candidate data that is aggregated at the program level. While these aggregate data might indicate overall program composition or quality in ways useful for accountability determinations, EPPs often need individual-level data on candidates to generate meaningful insights about specific program changes that might contribute to improvement.

The theory of action underlying current accountability policies also assumes that EPPs have the necessary capacity to use data for improvement. While a growing research base documents promising practices for building capacity within EPPs to use data for improvement (see Peck & Davis, this volume, Chapter 4), and organizations like Deans for Impact and others cited in this volume are supporting EPPs to grow this capacity, significant gaps remain across the field (Mandinach & Gummer, 2016). Data literacy itself is a challenge, with many EPPs struggling to help faculty and program leaders develop the skills necessary to access, interpret, and act on data for improvement (Mandinach & Gummer, 2013). Our own work suggests that creating the appropriate organizational scaffolds and structures to support data use for improvement in EPPs remains a challenge. Faculty and staff face many competing demands on their time, and existing incentive structures often do not reward them for engaging in the work of program improvement.

Finally, current accountability policies assume that EPPs—if given the right data to identify areas for growth—will know what to change about programs to generate improvements in candidate preparedness. But existing research provides inconsistent evidence about the features of EPPs that matter, for whom, and why (National Research Council, 2010). At the same time, there is a growing consensus that generating insights into the kinds of changes that hold promise for improving educator preparation will require substantially more robust efforts to collect and use common data on program design and content (Wilson, Floden, & Ferrini-Mundy, 2001; National Research Council, 2010; Data Quality Campaign, 2017; Southern Regional Educational Board, 2017). Common data within educator preparation—or data on candidate knowledge and skills and program performance generated using common measures and data collection protocols to enable cross-institutional comparisons—are particularly critical if EPPs are to learn from one another about promising practices (McDiarmid & Caprino, 2017).

These realities suggest that current accountability policies alone—even with more credible data at the center—are likely not sufficient to encourage necessary improvements in EPP quality. Thus, in 2016, leaders from diverse EPPs joined together to develop an approach to collecting and using data that would support meaningful improvement in educator preparation. These leaders are united by a belief that systemic improvement of educator preparation requires collecting data explicitly for improvement purposes and that provides actionable insights into the features of EPPs that matter,
while simultaneously growing the capacity of faculty and program leaders to use these data for program improvement. To actualize this belief, these leaders, in partnership with Deans for Impact, an education nonprofit that supports and empowers leaders in educator preparation, came together to develop a Common Indicators System (CIS).

The CIS is a network of diverse EPPs working to gather credible evidence on candidate knowledge and skill and program performance using common measures and data collection protocols and engaging in cross-institutional analysis of these data to collectively investigate how to better prepare future teachers. Common data collected through the CIS allow faculty and program leaders to learn with and from one another as they seek to improve their own programs and contribute to building a more robust evidence base about the features of teacher preparation that matter, for whom, and why. Through participation in the CIS, faculty and program leaders at participating EPPs also receive tools and resources to support shared inquiry into common data as well as targeted support to facilitate improvement planning grounded in these data that is intended help grow their capacity to use data for program improvement.

The CIS network’s cross-institutional design is unique in that it requires faculty and program leaders to negotiate both the context of their own institutions and the contexts of other participating institutions to collect data for shared inquiry. This type of data use requires new, more collaborative ways of working, both within and across EPPs. Initial development of the CIS occurred during the 2016–2017 academic year and involved work that occurred largely outside of participating institutions, with representatives from each EPP meeting at off-site convenings. This gave small teams of faculty and program leaders the opportunity, both individually and as a group, to work through the implications, challenges, and opportunities the CIS posed and to build a sense of momentum and collective action around the work. During the 2017–2018 academic year, the CIS moved into a new phase—the “pilot” year, during which CIS network EPPs agreed to pilot common measures and initiate a cycle of shared inquiry into CIS data. This required EPP representatives to negotiate the internal contexts of their respective institutions to lay a foundation for this new way of working and build awareness of the CIS among stakeholders who had not participated in its initial development while carrying forward momentum around the use of common measures to a broader group of program stakeholders.

Importantly, recent research on data use in educator preparation suggests that faculty engagement with certain kinds of common data (e.g., edTPA and program value-added ratings) can create “a shift towards more collaborative inquiry and shared responsibility for improvement” across EPPs (Davis & Peck, 2016). To date, however, the research on data use in educator preparation focuses primarily on the conditions that support data use within individual EPPs (e.g., Peck & McDonald, 2013, 2014; Peck, McDonald, & Davis, 2015; Davis & Peck, 2016), telling us little about how
EPPs might create the conditions necessary to collect common data for cross-institutional inquiry to inform program improvement. Thus, the work of CIS network participants during the pilot year creates an opportunity to examine the strategies institutionally diverse EPPs use to create conditions that support the collection and use of common data to accelerate program improvement.

Interviews with faculty and program leaders leading implementation of the CIS within their EPPs indicate that moving toward adoption of common measures for the purpose of cross-institutional learning and program improvement requires mobilizing faculty and other key stakeholders behind participation in the CIS as part of a broader improvement strategy and finding sufficient operational capacity to implement new systems and structures or to leverage existing ones to collect and use common data for improvement. Each EPP participating in the CIS network has charted a unique course to navigate this interacting set of challenges. However, their experiences do suggest some common strategies that can be tailored to local context to create the conditions for collecting common data to help inform program improvement.

First, to energize faculty and program leaders around participation in the CIS and to identify operational capacity, leaders should tap into stakeholders’ values and align a common-data effort to the existing work or projects of faculty. Next, leaders should carefully identify and empower champions to help shepherd the work and maximize the odds that faculty and program leaders are listening with receptive and open minds. Third, leaders should align the work to local priorities and help stakeholders see the connection between common data and larger institutional priorities so they see value in pursuing the effort. Finally, when it comes to the introduction of new common measures, it helps to start small and let the evidence do the talking: using early evidence from a smaller-scale pilot to demonstrate the value of common data for program improvement can be a powerful tool to move the work forward.

It is worth noting that these four strategies are not an exhaustive list, and indeed, there are many other approaches relevant for leaders who wish to initiate significant change efforts related to data use within their program. However, we believe these strategies provide a starting place for leaders at all levels and are easily customized to meet specific needs or augmented by additional change-management strategies. In this chapter, we will explore these four strategies in more depth, drawing on evidence and examples from eight EPPs participating in the CIS network to inform the efforts of other EPPs who want to advance the collection of common data to accelerate program improvement and promote field-wide learning.

**The Common Indicators System Network**

The CIS network includes eight public, three private, and two hybrid EPPs serving approximately 9,000 teacher candidates annually in undergraduate,
graduate, residency, and alternative pathways. These EPPs are united by a shared commitment to gathering formative data on candidate knowledge and skills and program performance using common measures and data collection protocols and engaging in cross-institutional learning informed by these data to identify promising practices for preparing future teachers. Since August 2016, this network has joined together approximately 60 faculty and program leaders—including deans, directors of assessment and teacher education, and clinical and tenure-track faculty—who have worked in partnership with Deans for Impact to co-construct the CIS in three distinct phases: Phase 1, which took place from August to December 2016, focused on identifying categories of data for a common system; Phase 2, which happened from January to June 2017, was centered on identifying instruments to collect common data in each of the categories selected in the previous phase; and Phase 3, which occurred during academic year 2017–2018, was the “pilot” year during which participating EPPs are testing the processes and procedures to collect common data.

Over the course of the first two phases of development, representatives from each participating EPP—known as data leads—met four times to discuss, debate, and ultimately decide upon the categories of data and the common instruments. The data categories identified emerged from a review of the literature, survey of member deans and program data leads, and consultations with technical advisors. They were selected and approved because stakeholders believed common data in the identified areas—observation of candidate instructional skill; assessment of candidate teaching beliefs and mind-sets; graduate feedback; and employer feedback—would be valuable for program improvement and informing field-wide learning.

A literature review and discussions with experts in the field then yielded 66 instruments that could be used to track teacher-candidate development and program performance in the identified data categories. Each of these instruments was subject to a content analysis based on selection criteria agreed to by leaders from within the CIS network. CIS network members then deliberated on a subset of the instruments that best met the group’s selection criteria. This process prioritized tools with evidence of validity and reliability and that could be implemented in a meaningful way across diverse program contexts. Based on these deliberations, network members agreed upon a set of common measures—drawn from the literature, and that have scientific bases as indicators or correlates of effective teaching—to gather data for the CIS (see Table 5.1).

During the 2017–2018 academic year, development of the CIS moved into its third phase—the “pilot” year—during which the 13 institutionally diverse EPPs that make up the CIS network agreed to pilot these common measures in their programs and engage in cross-institutional inquiry to inform program improvement. Each CIS network EPP identified a “trailblazing team” composed of three to five faculty and program leaders to lead implementation within their program. During the pilot year, trailblazing teams came together at a cross-institutional convening in early spring to
reflect on CIS implementation processes and procedures and begin planning for how they might use CIS data to inform improvement in their programs. Additionally, each trailblazing team had one representative on the CIS Advisory Panel for the pilot year. These representatives attended monthly calls to provide ongoing feedback on the development and implementation of processes and procedures to support data collection across the network.

As the primary goal of the pilot year was to test processes and procedures for collecting, storing, sharing, and collectively inquiring into data from the CIS, participating EPPs determined the size and scope of their pilot implementation; for example, some participating EPPs only implemented some of the common instruments, while others collected data from a subset of their teacher candidates. Data collection for the pilot year began during the 2018 spring semester, so the first opportunity that CIS participants will have to convene to review and reflect on common data was in August 2018. After the pilot year, the intent is to further refine the systems, structures, and processes of the CIS based on lessons learned from initial implementation.

Creating Conditions for Collecting Common Data

To better understand how diverse EPPs create the conditions for collecting common data to inform program improvement, we identified a sample of eight EPPs participating in the CIS network and conducted interviews with trailblazing team members from each program. These featured EPPs were selected because of their diversity in terms of size, institutional type, program pathways offered, and state context (see Table 5.2). They are not

<table>
<thead>
<tr>
<th>Data Category</th>
<th>Instrument Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation of Candidate</td>
<td>Assessed using the Classroom Assessment Scoring System (CLASS) observation tool, which captures evidence of the quality of teacher-student interactions</td>
</tr>
<tr>
<td>Instructional Skill</td>
<td></td>
</tr>
<tr>
<td>Teaching Beliefs and Mindsets</td>
<td>Assessed using a Teaching Beliefs and Mindsets Survey (TBMS) that combines the Short Teacher Sense of Efficacy Scale, Short Grit Scale, and items from the Culturally Responsive Teacher Self-Efficacy Scale</td>
</tr>
<tr>
<td>Graduate Feedback</td>
<td>Assessed using a modified version of the University of North Carolina-General Administration Beginning Teacher Survey (BTS), which captures graduate feedback on the quality of their preparation</td>
</tr>
<tr>
<td>Employer Feedback</td>
<td>Assessed using a modified version of the Massachusetts Hiring Principal Survey, known in the CIS network as the Employer Survey (ES), and which captures employer perceptions of the quality of a graduate’s preparation</td>
</tr>
</tbody>
</table>
necessarily representative of the entire CIS network but do represent diverse institutional profiles typical of those across the field of educator preparation more broadly.

In addition to selecting for institutional diversity, we featured EPPs that were diverse in terms of the common measures they are implementing during the CIS pilot and their approach to implementation (see Table 5.3). This enabled us to better understand the extent to which variation in the common measures adopted during the pilot might warrant different strategies for navigating implementation challenges related to collecting common data.

Table 5.2 Institutional Profiles of Featured EPPs

<table>
<thead>
<tr>
<th>Featured EPP</th>
<th>Enrollment(^a)</th>
<th>Public/Private</th>
<th>Pathways(^b)</th>
<th>Location(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relay Graduate School of Education</td>
<td>2,989 Private</td>
<td>Alternative, IHE-based; Graduate</td>
<td>Nationwide</td>
<td></td>
</tr>
<tr>
<td>Southern Methodist University</td>
<td>219 Private</td>
<td>Traditional &amp; Alternative, IHE-based; Undergraduate and Graduate</td>
<td>Dallas, Texas</td>
<td></td>
</tr>
<tr>
<td>Temple University</td>
<td>1,330 Semi-public</td>
<td>Traditional &amp; Alternative, IHE-based; Undergraduate and Graduate</td>
<td>Philadelphia, Pennsylvania</td>
<td></td>
</tr>
<tr>
<td>University of Nevada–Reno</td>
<td>887 Public</td>
<td>Traditional, Undergraduate, and Graduate</td>
<td>Reno, Nevada</td>
<td></td>
</tr>
<tr>
<td>University of North Carolina at Charlotte</td>
<td>1,492 Public</td>
<td>Traditional &amp; Alternative, IHE-based; Undergraduate and Graduate</td>
<td>Charlotte, North Carolina</td>
<td></td>
</tr>
<tr>
<td>University of Texas Rio Grande Valley</td>
<td>381 Public</td>
<td>Traditional; Undergraduate</td>
<td>Edinburg, Texas</td>
<td></td>
</tr>
<tr>
<td>University of Southern California</td>
<td>733(^d) Private</td>
<td>Traditional; Graduate</td>
<td>Los Angeles, California</td>
<td></td>
</tr>
<tr>
<td>University of Virginia</td>
<td>450 Public</td>
<td>Traditional; Graduate</td>
<td>Charlottesville, Virginia</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Total enrollment across all teacher-preparation programs, provided by the institution for SY 2017–2018.
\(^b\) Teacher-preparation program pathways offered, provided by the institution for SY 2017–2018.
\(^c\) Relay has 15 campuses in 10 states (LA, IL, CT, TX, DE, CO, TN, NY, NJ, PA) and Washington, DC.
\(^d\) USC’s total enrollment is reflective of Fall 2017–Spring 2018.
Table 5.3 Characteristics of CIS Pilot Implementation for Featured EPPs

<table>
<thead>
<tr>
<th>Featured EPP</th>
<th>CIS Pilot Measures</th>
<th>Implementation Approach</th>
<th>Trailblazing Team Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relay Graduate School of Education</td>
<td>CLASS, TBMS, BTS</td>
<td>Administering CLASS to a subset of candidates on the NY campus; Administering TBMS, BTS across all 15 campuses</td>
<td>*Campus Dean (3) *Chief Research Officer Assistant Dean Director of Research Assistant Professor</td>
</tr>
<tr>
<td>Southern Methodist University</td>
<td>CLASS, TBMS, BTS, ES</td>
<td>Administering CLASS and TBMS to a subset of candidates in the alternative program; Administering BTS and ES across all programs</td>
<td>*Associate Dean *Assistant Dean Academic Tech Dir. Assistant Professor Clinical Professor</td>
</tr>
<tr>
<td>Temple University</td>
<td>CLASS, TBMS, BTS, ES</td>
<td>Administering CLASS and TBMS to a subset of candidates in their elementary programs; Administering BTS and ES across all programs</td>
<td>*Assistant Dean Associate Professor Assistant Professor</td>
</tr>
<tr>
<td>University of Nevada-Reno</td>
<td>CLASS, TBMS, BTS, ES</td>
<td>Administering CLASS and TBMS across all programs</td>
<td>*Associate Dean Associate Professor Assistant Professor</td>
</tr>
<tr>
<td>University of North Carolina at Charlotte</td>
<td>CLASS, TBMS</td>
<td>Administering CLASS and TBMS across all programs</td>
<td>*Associate Dean *Director of Assessment Supervisor/Coach Associate Professor Assistant Professor</td>
</tr>
<tr>
<td>University of Texas Rio Grande Valley</td>
<td>TBMS, BTS</td>
<td>Administering TBMS and BTS across all programs</td>
<td>*Associate Dean Assessment Coordinator Professor Associate Professor</td>
</tr>
</tbody>
</table>
For each featured EPP, we conducted hour-long, structured interviews with trailblazing team members to discuss CIS implementation, focusing on the strategies they have used to address implementation challenges. We interviewed 13 trailblazing team members across the eight institutions. In all cases, we spoke with the trailblazing team lead(s) at each institution and in some cases were also able to interview an additional team member (see Table 5.3). Interviewees included associate and assistant deans, clinical professors, directors of assessment or heads of research, tenure-track faculty members, and other program staff. Interviews occurred in January and February of 2018, just before data collection began at most featured EPPs. By this point in time, each featured EPP had started to put final structures and systems in place to collect data on the common indicators for the pilot. Those adopting the CLASS observation tool, for example, were completing observer training and beginning to collect observation data, while those implementing one or more of the CIS surveys were preparing for administration. Importantly, however, none of the CIS network EPPs had begun using CIS data to inform improvement; for this reason, our focus was on understanding how trailblazers were creating the conditions to collect common data for cross-institutional inquiry and program improvement.

By design, our interviews yielded a variety of perspectives. Not all trailblazers we interviewed reported using each of the highlighted strategies to create the conditions for collecting common data, as each participating EPP has followed a unique roadmap. However, programs did utilize several common strategies regardless of their approach to implementation of the CIS pilot. In this chapter, we highlight under each strategy the institution(s)
that shared the most relevant experiences and provided the richest evidence. Taken together, the experiences of trailblazers we spoke to suggest that regardless of context, there are common strategies that can help any EPP interested in taking concrete action to create the conditions to collect common data for cross-institutional learning to inform program improvement.

**Tap into Stakeholders’ Values**

Trailblazing teams almost universally tried to tap into faculty and program leaders’ intrinsic motivations by helping them to understand the connection between the CIS and their own work or values. Extant research suggests that aligning efforts to use data for program improvement to local values and goals of faculty and program leaders is critical for enabling data use within EPPs (Cochran-Smith & the Boston College Evidence Team, 2009; Peck & McDonald, 2013, 2014). The experiences of trailblazers suggest that when it comes to common data, connecting the dots in this way for faculty can also help engage necessary stakeholders in efforts to collect new data using common measures in preparation for cross-institutional inquiry.

**Relay Graduate School of Education**

Relay Graduate School of Education is a fully accredited graduate school, which launched in 2007 as Teacher U. Relay now serves 2,989 graduate students in its MAT and residency programs on 15 campuses across the U.S. For the CIS pilot, Relay is administering the Teaching Beliefs and Mindsets Survey and Beginning Teacher Survey to candidates and/or recent graduates at all 15 of its campuses and is implementing CLASS with a subset of candidates on its New York campus. Relay’s trailblazing team includes two campus deans, a campus assistant dean, the director of research, and an assistant professor, as well as the organization’s chief research officer and the campus dean for Relay New York, who serve as team leads. Relay considered two main factors in composing its trailblazing team. First, it selected team members from the Relay campuses with the longest histories and the largest enrollments. Second, Relay identified multiple representatives from its research team as trailblazers. Its research team oversees all institutional data, so it was important to have perspective from that team on how to incorporate new data collection into existing practices and how to ensure that data shared are reported consistently across Relay’s campuses.

Relay has generated investment and excitement among faculty around the CIS in part by aligning the effort to faculty’s daily work. As the campus dean of Relay New York noted,

> In terms of generating buy-in from faculty, it is really the extent to which they believe that it is aligned to their work... So the things I think about [are], “Where can I find alignment to our curriculum or to our
vision for advisement?” And when I can tie the data to their individual work or their graduate students we can see a lot of momentum, and there is direct alignment [with the CIS].

Relay already assesses applicants on “grit” (perseverance and persistence toward long-term goals)—one of the constructs assessed by the Teaching Beliefs and Mindsets Survey—and teaches its graduate students to assess grit and self-efficacy among their own K–12 students. According to the campus dean of Relay New York,

We already have buy-in from our faculty that [grit] is an important measure because we teach it to our teachers so it’s a natural next step for us to begin to look at that data for our own graduate students.

Relay’s trailblazing team leads also underscored the importance of making an explicit connection for faculty between the CIS and the opportunity to access comparative observational data via the CLASS instrument. As the campus dean of Relay New York noted,

We [Relay] are at a place where we are really interested in comparative data as an organization and so is our faculty, so in talking about CLASS all I had to say is, “This is a tool that has been externally validated and we are part of this group with Deans for Impact, and [through the CIS] we are going to get comparative data.”

Tapping into stakeholders’ values and aligning the CIS to the core of their daily work has helped Relay build a sense of ownership and investment in CIS participation among faculty. With initial data rolling in, excitement is building about the opportunity to look at CIS data, which the chief research officer notes is “a big win” for engaging faculty in this work.

University of Virginia

The University of Virginia is a public institution located in Charlottesville, Virginia. Its Curry School of Education serves about 450 teacher candidates annually in graduate programs. The Curry trailblazing team is led by the director of teacher education and includes the director of assessment, accreditation, and accountability, an assistant professor, and a professor. These individuals were chosen for the trailblazing team because of their expertise in teacher preparation, teacher quality, and assessment. During the CIS pilot, Curry is implementing all four of the common indicators, including: CLASS, the Teaching Beliefs and Mindsets Survey, the Beginning Teacher Survey, and the Employer Survey. Curry was already collecting CLASS and variations of the Teaching Beliefs and Mindsets data prior to the CIS, so the direct alignment between that work and the CIS has made
the mechanics of implementation less challenging at Curry than for other participating EPPs.

However, when it came to mobilizing faculty to engage with the CIS, trailblazing team members at Curry similarly noted the utility of aligning implementation to a core value shared by faculty: the importance of learning from others to improve their programs and inform the field more broadly. In the early 2000s, for example, Curry participated in Teachers for a New Era, a multi-institution effort to improve university-based teacher preparation partly through the collection and analysis of more robust evidence (Carnegie Corporation of New York, 2001). This core value of improving teacher preparation through high-quality evidence and cross-institutional learning has permeated the Curry context (McDiarmid & Caprino, 2017). As one assistant professor noted,

We always want to get better and learn from other people and other approaches and so for us [the faculty], the value of being part of a network like DFI and this kind of endeavor of building indicators that we can look at across programs, that’s the real value-added . . . being able to learn from other places . . . and building the knowledge base of the field more broadly.

The desire to leverage common data to learn from other EPPs is layered on top of a well-established data culture at Curry in which faculty regularly spend time interpreting and making sense of data to inform program improvement. In the school’s data committee and teacher education meetings, for example, faculty and staff regularly spend time thinking about what the data they already collect tells them and how they can collect additional data to gain deeper insight into candidate progress. In leveraging this existing data culture and aligning implementation with a core value shared by faculty, the Curry trailblazing team has been able to cultivate broad support among its stakeholders for collecting common data as part of the CIS.

Identify and Empower Champions

The importance of leadership in promoting data use activities is well documented in the extant research on data use within EPPs (Sloan, 2013; Davis & Peck, 2016). Our interviews with trailblazing teams suggest that they too were acutely aware of the importance of identifying the right messengers when it came to implementing the CIS. Almost every trailblazing team we spoke with made strategic decisions about everything from the composition of their core trailblazing team, to who should be the “face” of the work, to whom to tap as CLASS trainers. Empowering the “right” champions to serve as leaders for initial and ongoing implementation has helped trailblazing teams cultivate genuine interest in the idea of common data among faculty and other stakeholders and to identify the operational capacity necessary to coordinate data collection using new measures.
University of Nevada–Reno

The University of Nevada–Reno is a public institution whose College of Education serves roughly 900 teacher candidates in traditional graduate and undergraduate programs. During the CIS pilot, the college is piloting all four CIS instruments, including: CLASS, the Teaching Beliefs and Mindsets Survey, Beginning Teacher Survey, and Employer Survey. The College of Education’s CIS implementation is being led by a three-person trailblazing team that is headed by the associate dean and includes an assistant professor and an associate professor who is also a program chair.

From the start, the associate dean positioned herself as a highly visible champion of the CIS who was also intimately involved in the project’s execution—from leading implementation, to communicating to faculty, to managing the transfer of data. Given her position within the college, the associate dean’s level of involvement sent a strong signal about the project’s importance and ensured that she was familiar with every detail of implementation.

The college also strategically selected its trailblazing team, tapping full-time academic faculty who were deeply involved in educator preparation, who were excited about the project and the work involved, and who had some positional authority. This created a core team that combined visibility and authority at the college level with direct influence at the program level.

While the college was purposeful in identifying the initial group of “champions,” one happy accident might have created its most powerful ambassador. One trailblazing team member who had been tapped to become a CLASS trainer returned to campus after training feeling ambivalent about the instrument. Yet she stayed committed, completed extra training videos, and passed the reliability exam for both upper elementary and secondary grade levels. Through that process, she was converted from a skeptic to a proponent. At the end of a CLASS training held at the college, this faculty member stood and shared her story—the doubts she overcame and her belief in the tool. It was a powerful moment, recalled the associate dean,

> It’s so much cognitive load, and you’re overwhelmed and unsure of your own abilities—having someone who has gone through that process is really helpful. Having someone who can speak to . . . any lingering doubts is powerful.

University of North Carolina at Charlotte

The importance of selecting and empowering the right leaders for the work was echoed by the trailblazing team at the University of North Carolina at Charlotte’s Cato College of Education. The University of North Carolina at Charlotte (UNCC) is a large, public institution whose College of Education serves about 1,400 candidates in both undergraduate and graduate programs. The College of Education’s five-person trailblazing team was
intentionally selected to include broad representation from faculty and program personnel with broad knowledge of and engagement with assessment practices within the college. Led by the associate dean, the trailblazing team also includes an assistant professor, an associate professor, a university supervisor, and the director of assessment and accreditation.

According to the director of assessment and accreditation, the choice of trailblazing team lead was particularly important for UNCC,

Having the associate dean as the messenger and face of this work has been incredibly helpful and that has gone a long way because she is highly respected.

Thinking strategically about other messengers, including who to select as CLASS trainers, also proved important for UNCC. As the director of assessment and accreditation continued,

We picked a trainer from our office of field experiences, and then we chose someone who has a ton of background on CLASS, and then someone that everyone loves and is a veteran, and so those three together were a really important strategy to make this less of a challenge.

Both the trailblazing team members from UNCC with whom we spoke recommended that other institutions considering adopting common measures select their champions carefully to help mitigate resistance from stakeholders to new measures or procedures.

Align to Local Priorities

The importance of context in shaping the process of using data for improvement is well documented in the existing research on data use generally (e.g., Coburn & Turner, 2011), and within EPPs specifically (Davis & Peck, 2016; McDiarmid & Caprino, 2017). When designing implementation plans to collect common data, trailblazers also recognized the importance of considering existing contextual factors, including accreditation requirements, state policy context, and historical or ongoing university or college initiatives. By demonstrating alignment between local priorities and the CIS, trailblazers leveraged their context to “make the case” to faculty and program leaders about the broader value of participating in the CIS. Across all the participating EPPs we spoke with, aligning data collection to local priorities helped increase investment in collecting common data through the CIS.

Southern Methodist University

Southern Methodist University is a small, private university located in Dallas, Texas. The university’s Simmons School of Education and Human
Moving Toward Common Measures

Development (Simmons) currently serves about 220 teacher candidates and is implementing all four CIS measures during the pilot year, including: CLASS, the Teaching Beliefs and Mindsets Survey, Beginning Teacher Survey, and the Employer Survey. Coordinating implementation is a five-person trailblazing team, which includes an assistant professor, a clinical professor who is also the director of the Teach for America program, and a data systems administrator, as well as an associate dean and an assistant dean who serve as team leads for Simmons. To identify the trailblazing team, Simmons relied on volunteers who were interested in the work around common indicators and looking more closely at the data that were being collected.

At Simmons, the CIS fit naturally into an ongoing effort to build a strong culture of evidence. The assistant dean and associate dean joined Simmons in 2015 and 2016, respectively, and brought with them prior experiences with data use and assessment in higher education. Simmons’ new dean, who was appointed in August 2017, has also prioritized creating a culture of evidence and has served as an ambassador for the CIS work, consistently reminding faculty that the purpose of all data collection and analysis is to use evidence to improve.

Internal communication at Simmons about the CIS intentionally focused on the importance of evidence for improvement to reinforce the alignment of the CIS with this broader organizational priority. According to the associate dean,

This has never been about being good or bad or being marginal. This is always about what kind of data would be helpful to you, that you would want to know about teacher-candidates. . . . We can learn from our graduates and how well their students are performing.

The Simmons trailblazing team also aligned CIS participation to external factors that shape local priorities within the school of education. Under recently passed legislation, the Texas Education Agency (TEA) must provide educator preparation programs with data on their graduates. To do this, TEA is creating an “educator data dashboard” and redesigning statewide principal and beginning teacher surveys. TEA expects to pilot a new principal survey during the 2018–2019 school year and subsequently to revamp the beginning teacher survey.

Simmons’ involvement in the CIS network positions them to influence these external policy changes. As the associate dean remarked,

TEA has validated [that] we can be on the cutting edge of this work. We can be seen as leaders who are informing the state and educator preparation across the country. For faculty and for people who run programs, that’s music to their ears.

According to the associate dean, faculty were getting the message from all corners—not just from the trailblazing team—that the CIS was work
Simmons needed to be doing for the future. When reflecting on this, the associate dean remarked that this helped

create a bigger picture of where people are going instead of making it feel like an isolated effort that’s not tied to something greater and not connected to kids and schools and districts.

The University of Texas Rio Grande Valley

The University of Texas Rio Grande Valley (UTRGV) is a public university in the southern-most region of Texas. Its College of Education and P-16 Integration serves nearly 400 teacher candidates in undergraduate programs. The college’s trailblazing team includes four members: an assessment coordinator, an associate professor, a professor, and the associate dean for assessment and accreditation, who also serves as the trailblazing team lead. Like other participating EPPs, UTRGV selected their trailblazing team strategically. The team’s two faculty members were selected because they were founding members of the college assessment committee and are familiar with the quality assurance system in the college. In addition, the associate professor is program coordinator for the largest specialization area in the elementary teacher-preparation program, while the professor is the faculty lead for the college’s CAEP Standard 4 team, which is focusing on assessing program impact. According to the UTRGV team lead, the assessment coordinator’s role is to support the associate dean for assessment and accreditation in promoting continuous improvement and developing a culture of inquiry within the college; therefore, it was important to include them on the trailblazing team as well.

The college is piloting two CIS measures during the pilot: the Teaching Beliefs and Mindsets Survey and the Beginning Teacher Survey. These measures were strategically selected for the CIS pilot, in part because they aligned to current initiatives being led by two existing college committees: the Assessment Committee and a CAEP steering task force. To build excitement among participating faculty, the associate dean targeted initial communications about the CIS to these committees. According to the associate dean, participation in any new initiative “will always be a lot better received by the broader faculty when they see their peers excited about the project and really being champions for the implementation,” so aligning the CIS to existing priorities of faculty on these committees was an important first step to building investment in the work within the college.

The first of these committees, the Assessment Committee, is a standing committee tasked with overseeing all assessment activity within the college. The associate dean described the committee members as “forward-thinking” about the college’s assessment work. As the CIS was being developed, the Assessment Committee was independently developing a professional dispositions
inventory, which was intended to play a similar role as the Teaching Beliefs and Mindsets Survey plays within the CIS. According to the associate dean,

It was important for this group to understand what we’re doing and the purpose of also implementing the Teaching Beliefs and Mindsets Survey to have common data across institutions. Then we’ll see where this takes us. . . . We’ll let the data inform our work.

When presenting to the committee, the associate dean emphasized that the CIS was an opportunity for the committee to compare data collected from their internally developed instrument with data collected from the set of externally validated and reliable scales that make up the Teaching Beliefs and Mindsets Survey. The associate dean continued,

I first had to validate all of the work that the committee has done and present [the Teaching Beliefs and Mindsets Survey] as an additional source of information that should complement what we are doing, and not supplant or replace the efforts that have already taken place.

The CAEP steering task force, which was created to support the college’s accreditation review, was the second committee targeted for outreach. Within the task force, a lead faculty member was designated for each CAEP standard, including Standard 4, which is focused on a program’s impact on student learning. The faculty member assigned to Standard 4 formed a team of faculty to identify how to best gather evidence to meet this standard; all the participating faculty volunteered because of their interest in understanding whether (and how) the college’s graduates are influencing student learning. As the associate dean remarked,

For the Beginning Teacher Survey, this was the perfect group to go to for ownership and investment in the participation of this project. This team had just been formed, and was looking for how to meet the standard . . . so it fit perfectly in the work.

Now that the CIS has been presented to these two committees, the trailblazing team plans to continue to expand faculty awareness of and engagement in the project. For example, following the spring CIS convening, the different members of the assessment committee presented the CIS to their department faculty and are discussing how they might share CIS data with these faculty moving forward.

**Start Small and Let the Evidence Do the Talking**

Even after tapping stakeholders’ values, identifying and empowering champions, and aligning to local priorities, many trailblazing teams faced a
chicken-and-egg problem. They wanted to demonstrate through evidence how participation in the CIS network could enhance program practice and improve the teacher-candidate experience. But demonstrating the value of participation required participation itself. To solve this problem, many trailblazing teams elected to start small and use initial data to build the case for expanded participation.

Temple University

Temple University is a large, semi-public institution located in the heart of Philadelphia, Pennsylvania. Its College of Education serves more than 1,300 undergraduate and graduate teacher candidates in both traditional and alternative programs. The college’s trailblazing team includes an associate professor, an assistant professor, and the assistant dean of teacher education, who serves as the trailblazing team lead. Members of the trailblazing team were identified based on three factors: their experience successfully collecting and analyzing data; their experience providing leadership; and their membership in the college’s Center for Assessment, Evaluation, and Education Policy Analysis. During the CIS pilot, the college administered the CLASS observation tool and the Teaching Beliefs and Mindsets Survey to a subset of candidates in its elementary programs; it administered the Beginning Teacher and Employer surveys across all its programs.

The college’s recent history with change and innovation efforts informed the design of its CIS implementation plan. In the years prior to rolling out the CIS, the university mandated a Responsibility Center Management (RCM) budget model within the college, the implementation of which began with an intensive period of data gathering. College leadership held a retreat for faculty to review these data and to discuss potential recommendations for implementation. While not all faculty agreed with every aspect of the RCM implementation, they respected the role that data played in informing the college’s future direction. This experience highlighted the value that the college’s faculty place on evidence.

Other past initiatives, including one focused on cross-course redesign within teacher education, were “met with skepticism,” according to the college’s assistant dean. “Where’s the data on this?” college faculty would ask when faced with a new initiative.

Based on this history, the college’s trailblazing team believed conversations with faculty about the CIS would be more productive if they were anchored by real common data. For these reasons, the college chose to implement the CIS with a smaller number of teacher candidates and use structures—like video observations for CLASS scoring—that would minimize the impact to faculty work streams. Relative to other trailblazing teams, the college kept their pilot small and relied more heavily on existing members of the trailblazing team instead of expanding the work to other faculty and stakeholders. The trailblazing team hopes that this approach will generate data
they can use to open a conversation with faculty about broader engagement with the CIS.

University of Southern California

The University of Southern California is a large private university located in the heart of Los Angeles. Its Rossier School of Education (Rossier), enrolls about 700 teacher candidates annually in both online and on-campus programs. For the CIS pilot, Rossier implemented both the CLASS observation tool and the Teaching Beliefs and Mindsets Survey to a subset of candidates in their on-campus elementary and secondary programs. Implementation was coordinated by a four-person trailblazing team that includes the director of accreditation, a project specialist, an associate clinical professor and MAT program chair, and a clinical professor. Members of the trailblazing team were chosen based on their experience and commitment to the CIS effort.

Like Temple and many other EPPs participating in the CIS, Rossier faced a similar chicken-and-egg problem during the pilot year. Rossier had recently completed a redesign of its Master of Arts in Teaching (MAT) program to increase vertical and horizontal alignment. This redesign was spurred by a change to California standards for teacher-preparation programs, but program leaders were able to draw on a five-year external evaluation of the MAT’s efficacy to inform and affirm their vision for the redesign. The goal of the external evaluation was to assess the effectiveness of MAT graduates from the perspective of teacher candidates, graduates, mentor teachers, administrators, and employers. This experience affirmed the value Rossier and its faculty place on using robust evidence to inform programmatic design decisions.

Rossier leadership and the trailblazing team saw the CIS as a way to extend and deepen the work begun by the initial five-year evaluation and as a way to understand how the new program design was influencing teacher-candidate outcomes. But faculty, who were supportive in principle, were also stretched thin by initiative fatigue.

Rossier originally intended to pilot CLASS with a large number of its teacher candidates and involve its faculty in conducting the observations. However, acknowledging the many competing demands currently placed on faculty, the trailblazing team decided to reduce the number of candidates observed and rely on the trailblazing team to conduct observations. During the pilot year, instruments were implemented only with teacher candidates enrolled in Rossier’s “on-ground” program, based at the school’s Los Angeles campus. Once the trailblazing team has collected pilot data, they will launch an evidence-informed conversation with faculty about a broader implementation of the CIS across Rossier.

This extended rollout also created space for Rossier to identify other resources to support an expanded implementation. For example, the
trailblazing team plans to crosswalk the CLASS rubric with other rubrics being used by faculty to assess candidate progress during student teaching. “Drawing similarities between rubrics will help faculty feel like they’re not siloed, that they don’t need to put different hats on for each tool,” remarked the project specialist.

Conclusion

The adoption of common, cross-institutional indicators of candidate knowledge and skills and of program performance holds great promise for both practice and research in educator preparation—but moving from concept to reality is not without its challenges.

The experience of CIS trailblazers suggests that support at the dean level is critical. Each of the trailblazing institutions featured here is participating in the CIS, in part, because its dean is invested in the goal of enabling cross-institutional learning, grounded in high-quality evidence, to better inform program improvement and stimulate field-wide learning. These deans have created the time and space for trailblazing teams to prioritize and carry out this work and served as advocates for the work to the broader faculty. In the absence of this strong leadership from deans, it is unclear to us how EPPs might create the conditions necessary to collect and use common data to accelerate program improvement and inform field-wide learning.

The development and initial implementation of the CIS, however, also demonstrates that while strong dean leadership is an essential precondition for this work, it is not sufficient. Successfully moving toward the adoption of common measures requires the involvement of leaders at all levels within an educator preparation program, from associate deans to directors of teacher education to clinical faculty. Even with strong dean support, these faculty and program leaders face a range of challenges, including energizing their colleagues around collecting common data as part of a meaningful program improvement effort and identifying the operational capacity needed to successfully adopt new measures. Yet as evidenced by the experiences of the featured EPPs these challenges are surmountable.

In fact, despite the diversity of their program structures, size, state and local context, and approach to implementing the CIS measures, faculty, and program leaders from CIS network EPPs identified a handful of successful strategies that can help all EPPs successfully implement an effort to gather common data for program improvement. This work takes patience and persistence to keep everyone focused on a long-term goal. Tapping into stakeholders’ values, carefully identifying and empowering champions, aligning the work to local priorities, and starting small and letting the evidence do the talking are all promising strategies for EPPs to leverage on their journey to gather the common data needed to inform field-wide learning and improve teacher preparation.
Moving Toward Common Measures

It is important to note that at this stage of the CIS pilot it is hard to know how effective these strategies will be for ensuring continued support for and meaningful engagement with the CIS among a broader group of faculty and program leaders within participating EPPs. However, we believe the structure of the CIS work—which started small, with just a representative or two from each participating program, and gradually expanded to include broader representation from faculty and staff over the course of the pilot year—will contribute to its sustainability. The project is not “owned” by any one person at any institution but instead is a collective effort among teams within individual institutions and across the network as a whole. In the pilot year alone, we saw the addition of faculty and staff to trailblazing teams, and a broader representation of faculty and staff from many participating EPPs attending CIS network convenings. Additionally, we saw trailblazing teams continue to participate in the CIS network despite changes in dean leadership, because they see tremendous value in collecting common data for the purpose of cross-institutional inquiry. While we know that CIS network EPPs will likely face resistance to the CIS from some of their faculty and staff, our hope is that trailblazing teams will continue to engage their faculty and program leaders in this work in ways that make sense for their context and that doing so will help embed the CIS as an integral tool for program improvement across participating EPPs.

Finally, common data is not a panacea, and as emerging research demonstrates, how it is used to inform improvement (or not) matters significantly for its value to individual EPPs and the field more broadly (Davis & Peck, 2016; Peck & Davis, this volume). At this stage of the work, it is too soon to tell whether and how EPPs participating in the CIS network might leverage common data for improvement. But using common data for improvement first requires that EPPs commit to the challenging work of adopting and implementing common measures to gather the data in the first place. Trailblazers that we interviewed agreed that longstanding criticisms of U.S. teacher preparation need to be met with robust, empirical evidence that diverse EPPs can indeed work together to improve the way teachers are prepared and improve the educational opportunities provided to the children of this country. Common data can serve as the catalyst necessary for this work to unfold. The strategies leveraged by CIS network EPPs to create the conditions to support this work can help inform other EPPs working to carve their own path to move toward adoption and implementation of common measures.

Notes

1. Institutions participating in the CIS network include: Arizona State University, Boston Teacher Residency, Loyola Marymount University, Relay Graduate School of Education, Southern Methodist University, Temple University, Texas Tech University, University of Nevada–Reno, University of North Carolina–Charlotte,
University of Southern California, University of Texas Rio Grande Valley, University of Virginia, and Urban Teachers.

2. These 66 instruments represent a combination of those collected from the programs participating in the CIS as well as those surfaced through a literature review and from conversations with experts in the field. These instruments are by no means the only instruments used by EPPs across the field but represent a systematically gathered cross-section of instruments used to track teacher-candidate development and understand program performance.


References


Peck, C. A., & Davis, S. (this volume). Building capacity and commitment for data use in teacher education programs. In E. Mandinach & E. Gummer (Eds.), Data for


6 Connected for Improvement
The Teacher Preparation Data Model and TPP Dashboard

Ana Quintana, Bryan Richardson, and Cari Reddick

About this Chapter

The preparation and support of teachers is a shared responsibility among teacher preparation programs (TPPs), certification agencies, and those hiring program graduates. When everyone works together, the quality of the teaching profession is strengthened. To do their part and align their goals in a continuous improvement cycle, each of these groups needs access to better data at specific times in the teacher talent pipeline.

In an attempt to address the challenge of data availability in teacher preparation, starting in 2016, a partnership of funders (the Bill & Melinda Gates Foundation, the Michael & Susan Dell Foundation, and the Overdeck Family Foundation) and UPD Consulting initiated the Teacher Preparation Data Project (referred to in this chapter as “the project”). The goal of the project was to make the integration of and access to data critical to teacher preparers easier and less expensive and to make innovative solutions developed in one location scalable to others.

UPD conducted research, gathered requirements, and developed technology to create the Teacher Preparation Data Model (TPDM), involving key players as stakeholders in the process. Central to the TPDM is the use of a data and interoperability standard called Ed-Fi, which enjoys a relatively broad community of users in the K–12 market. TPDM is an extension to the Ed-Fi data standard for teacher preparation that will harness and integrate preservice and in-service data to be used to prepare teacher candidates and support the growth of current teachers. The TPDM extension enables comprehensive data aggregation over the span of a teacher’s entire career, from entry into a TPP, through knowledge and skills demonstrated in fieldwork experiences, to placement and performance as an in-service teacher. TPPs, state education agencies (SEAs), local education agencies (LEAs), and schools can access and review data through automated data connections to develop strategies to ensure teachers are prepared to meet the needs of their students.

UPD is also developing a dashboard and reporting tool aligned with the Ed-Fi TPDM to enable TPP stakeholders to have access to the right data at the right time to aid in decision making for program improvement. For this TPP dashboard project, UPD is also utilizing a collaborative process of
gathering requirements, feedback, and best practices from TPPs and partners. The dashboard tool will be made available through the Ed-Fi Exchange so that TPPs and state and district partners may choose to implement it out of the box or customize it to tailor it to their needs.

This chapter tells the story of the TPDM with three main goals.

1. To inform the teacher preparation space on what the TPDM is, how it works, and how it can address the challenges of data availability for teacher preparation practitioners.
2. To detail effective stakeholder engagement processes used to anchor the capabilities of the TPDM in real problems of practice.
3. To discuss the challenges and opportunities in the road ahead for the resources being developed in the TPDM as its focus turns to scaling shared solutions, like data dashboards, into a new community of users.

Background Context—Why This, Why Now

It’s hard to understand the need for the TPDM without first understanding the broader context of the technological, policy, and data use movements that came before it. These market foci combined to catalyze both the data technologies that underpin the TPDM and the ideological biases for where these technologies are focused.

Demand for Better Data and Data Use at TPPs

There is a general consensus that supporting and preparing teachers is a shared responsibility between TPPs, SEAs who certify graduates, and local schools who hire graduates. Each of these stakeholders has different goals for data usage, and all should be involved in the creation of tools that provide ongoing data-based decision making and program improvement leading to professional growth for the nation’s teachers.

For all of the innovation and excitement present in the teacher preparation space today, there is much work yet to be done. As reflected in research through the Measures of Effective Teaching (MET) Project (Bill & Melinda Gates Foundation, 2016), great teaching is the most important in-school factor in determining student achievement. The teaching profession faces current challenges of high teacher attrition rates and decreasing enrollment in TPPs. This creates a critical need to provide TPPs with data and tools so that they can effectively analyze how they are preparing their candidates to enter the profession as successful teachers.

TPPs may be required to collect effectiveness data so that states can monitor program quality. In addition, the Council for the Accreditation of Educator Preparation (CAEP) standards, affecting more than half of providers across the country, requires programs to meet more stringent performance metrics, including K–12 outcomes. CAEP standards require data to track where graduates teach and the performance of their students.
Although many programs are currently collecting large amounts of data, this data is not consistently being used to guide program decisions. Typically, TPP staff have to work across several internal and external systems to get the data needed for advising individual teacher candidates and for reviewing program performance, and they have limited resources to do so.

One of the most significant challenges is the ability for TPPs to access and review in-service data on their program graduates to analyze how effectively the program has prepared them. This involves partnerships with TPPs, states, and districts to collectively share and review data. Many states are not yet equipped to integrate and deliver this information to TPPs and districts. Many districts and states may be restricted from sharing this data, or the data exist in multiple systems, making it technologically challenging to share effectively. Cultural and policy changes are needed, as sharing certain data is often not the norm. In some cases, where data sharing agreements have been achieved, these are dependent on relationships formed among specific staff members rather than a comprehensive process.

These challenges impact many: the K–12 students teacher candidates are teaching, the school districts with whom TPPs partner, the teacher candidates in the program, the faculty who teach the courses, and the program administrators who are responsible for designing and improving the overall TPP experience. Figure 6.1 illustrates the critical relationships among TPPs, states, and districts.

**A Solution for K–12: Ed-Fi Data Standard**

Promisingly, however, the K–12 domain has done laudable work over the past five years to develop common data standards that lower the cost of data integration and make the innovations of one state or district transferable to
another. The Ed-Fi data standard is the widely adopted, open-source data standard developed by the educational community for the betterment of the community. The Ed-Fi data standard serves as the foundation for enabling interoperability among secure data systems and contains a unifying data model designed to capture the meaning and inherent structure in the most important information in the K–12 education enterprise.

The Michael and Susan Dell Foundation (2013) announced the Ed-Fi Alliance:

*A New Standard in Data-Driven Education*

The Ed-Fi solution addresses a persistent problem for teachers and administrators: accessing valuable data locked away in different data systems and tools that do not interoperate. The Ed-Fi solution accelerates student achievement by extracting student information from a variety of sources, and integrating the data into Web-based dashboards, reports and other applications that are, with the right permissions, accessible to educators and other parties on demand. This free educational data standard and tool suite, developed with input from teachers, school leaders, local and state education agency officials, and vendors nationwide, is composed of:

- a unifying data model
- a data exchange and application framework
- sample dashboard source code

Launched in July 2011 to improve K–12 student achievement, the Ed-Fi solution also enables school and district administrators, and state and federal agencies to supplement or replace tools used for broader state or federal accountability reporting purposes.

In particular, the Ed-Fi data integration toolset (www.ed-fi.org) has emerged as one of the most promising open-source data integration platforms. Since its inception, Ed-Fi has been licensed by more than 100 education organizations (states and school districts) (source: www.ed-fi.org/what-is-ed-fi/current-licensees/). Ed-Fi’s largest contribution, and the main reason for its success, is that it is (1) open source, (2) a national standard, and (3) addresses a deep and broad set of national needs.

As a data model, Ed-Fi integrates an exhaustive list of data elements used in K–12 education within one integrated logical mode. It is designed to accommodate information from all the source systems used to do business in education and to create “a single version of the truth.” The model itself is comprised of more than 1,600 data elements in domains ranging from student attendance to finance to assessments. Whereas the Ed-Fi data model articulates how data can be integrated conceptually, the Operational Data Store (ODS) provides the technology developers will use to operationalize the model physically.
This Ed-Fi data model allows for states and school districts to think of their data as one enterprise asset rather than as an amalgam of several disconnected systems. This enterprise perspective is powerful. First, by starting with a single integrated vision of all the data education organizations need, it is not as disruptive when organizations inevitably need to change source systems. Second, by having all of their data in one place, organizations are more nimble in being able to answer the questions that educators need to do their work in educating students. And third, by standardizing the way that data is organized—and making the model free—it is easier, faster, and cheaper to share technology built on the standard.

As we will further describe later in this chapter, the TPDM was developed as a set of extensions to the Ed-Fi K–12 core data standard. The goal in creating the TPDM is to establish this same open foundation for data integration for teacher preparation. This data model was principally designed around the needs of TPPs, but it will have substantial value for school districts and state education agencies in pulling together data that span the full spectrum of a teacher’s career.

To create the model, we aligned to the existing Ed-Fi data model by making use of the common data elements between K–12 and TPPs. In practice, this involved starting with the list of data elements needed by teacher preparation and comparing them to the data elements in the Ed-Fi K-12 data model. Where there was a match (e.g., student ID is used for a K–12 student and is also needed for a teacher candidate), we made use of that data element in the TPDM. Where there was not a match to a data element in the Ed-Fi K–12 data model, we noted these for extensions that would be created. Leveraging the existing Ed-Fi data standard should reduce the cost and difficulty of integrating teacher preparation data among the many states who are already Ed-Fi licensees, and hopefully it will accelerate innovation as both Ed-Fi and TPDM improve in the future.

In summary, the demands for TPP data and advances in data integration technology in K–12 set the stage to bring focus and energy to improving the use of data to measure the performance of to help improve TPPs.

How the TPDM Idea Evolved

In addition to the market forces described in the previous section, there were a series of projects that UPD Consulting worked on that helped to inspire us to formulate the plan and strategy for the TPDM.

EdPrepStat with the Colleges of Teacher Preparation in the State of Hawaii

UPD worked with the seven major colleges of teacher preparation in the State of Hawai‘i under the leadership of the University of Hawai‘i at Manoa to implement a performance management routine called EdPrepStat. This
Connected for Improvement

process engaged the colleges of education in rigorous data collection and analysis, program improvement planning, and performance management routines over 18 months. In this project, UPD supported program faculty in the development of goals and theories of action and developed metrics to be tracked to measure progress toward their goals. The project also involved structured capacity building so that participating programs are positioned to continue EdPrepStat once UPD’s support was complete.

Shelby County Schools Talent Management

UPD worked with Shelby County Schools (Memphis, Tennessee) to develop a data integration and dashboard within the Ed-Fi platform based on the findings from the Strategic Talent Management Decisions for Principals work at Vanderbilt University (2014). The study identified how principals use teacher effectiveness data to make talent management decisions. Based on the findings of this study, the dashboard was designed to include measures to inform hiring, placement, evaluation support, and leadership.

As this work progressed, UPD received intense interest by TPPs to use these Ed-Fi talent management dashboards as the pathway to provide the information they need to understand the impact of their graduates and to more strategically improve their programs. When TPPs saw the tool and imagined it describing their preservice residents, they saw what they were looking for.

Arizona State University’s Building Blocks

The Mary Lou Fulton Teachers College (MLFTC) at Arizona State University (http://education.asu.edu) is widely recognized as one of the most innovative educator preparation programs in the country. Over the course of the last four years, MLFTC has developed a data reporting platform that is integral to the success of its iTeachAZ program, a rigorous model for teacher preparation that includes an intensive clinical experience, regular feedback, and performance-based assessments. The success of this program at scale hinges on the collection and consistent use of data about teacher candidates, made readily available to the many people involved in supporting them. The iTeachAZ data platform is an integrated dashboard application that provides teacher candidates, mentor teachers, clinical faculty, and administrators with detailed information on the progress and unique needs of MLFTC students. The platform has been presented nationally and is broadly recognized as one of the best data tools in teacher preparation across the country. While powerful, ASU has struggled to make iTeachAZ scalable and open to sharing with the more than 100 interested TPPs across the country.

Starting in the summer of 2015, UPD began a working relationship with the MLFTC to better understand how data can support teacher preparation
pre-service. The work with MLFTC led to conversations with hundreds of TPPs across the country on the idea of leveraging Ed-Fi to support teacher preparation. As these conversations evolved, so many TPPs saw the value of a tool like iTeachAZ but also realized the barriers to implement it within their own organization because it was developed on a proprietary tool and designed specifically for MLFTC. This is how the need and value of using a standardized, open-source solution like Ed-Fi became apparent. Through these discussions and realizations, TPPs saw the TPDM as a critical opportunity to advance as a field and to do so in a way that prioritizes the values TPPs hold dear—using data for continuous improvement and for collaboration.

Early Thoughts about Extending Ed-Fi for Teacher Preparation

While gathering advice on the process for creating the extensions for teacher preparation using Ed-Fi technology, one point stood out: users can’t believe that they are being given something that is already the best there is. They have to believe that they are creating something together to serve a set of national needs and are doing so with the best partners there are.

This is a subtle but critical distinction. The solution for teacher preparation could not simply be the repurposing of existing technology that TPPs like. It needed to be the product of an authentic stakeholder engagement process that positions early adopters as co-creators and founding partners for something that does not exist today.

Of the many lessons we learned about teacher preparation during the investigative stage, one has stood out. We are not the experts in using data for teacher preparation. The practitioners working in programs, SEAs, and LEAs are the experts. As such, our first role in this project was to create a process that listens to and channels the goals of stakeholders, trusting that they know how to articulate what they need. Once that process unpacked their needs, our role was to translate their vision into the Ed-Fi technology.

The Teacher Preparation Data Project

June 2016 marked the official start of the grant-funded Teacher Preparation Data Project, which set out to accomplish the following goals:

- Clearly define the data needs of TPPs to strengthen their programs and build cultures of continuous improvement.
- Build an organized inventory of data elements that will be needed to empower TPPs, districts, and states with answers to the key programmatic questions they seek to answer.
- Empower teacher preparation stakeholders with data to make timely, relevant, and impactful improvement decisions.

TPPs, both traditional and alternative, require data to understand: (1) who they should be recruiting; (2) how well their teacher candidates demonstrate
the knowledge and skills needed to be successful teachers; (3) where their
teachers are placed and how they perform once there; and (4) what their
impact is on K–12 students. TPPs also face many of the same challenges as
K–12 districts in that data exist in disparate data systems or may be tracked
manually in spreadsheets. In addition, there are many challenges to enable
states to integrate and deliver this information to TPPs and districts.

UPD believes that the foundation to a better teacher preparation model is
ease of access to powerful, interconnected data. Data should empower each
stakeholder in the teacher preparation process to make timely, relevant, and
impactful decisions so they can do their work more effectively. Data should
also help stakeholders identify the preservice supports provided in programs
that result in teacher effectiveness in the classroom. These data need to be in
one place in an easy-to-understand format so that stakeholders can use this
information first and foremost toward helping every program better prepare
teachers for their classrooms. We kept these big ideas and objectives in the
forefront through the process of creating the open-source, nationally scal-
able data solution that allows for programs, states, and districts to better
visualize and use their preservice data and connect it with in-service data on
practicing teachers and their impact in the classroom.

**Stakeholder Engagement: A Collaborative Approach**

Stakeholder engagement in this project was quite different and more chal-
lenging than a traditional requirement-gathering process: to scope and
design a data model and Ed-Fi technology that largely does not yet exist.

In this context, stakeholder engagement put a premium on precision
around what the user wants but also being open and accommodating on
the pathways to get there. UPD employed the tools of human-centered de-
sign and design thinking, among several other frameworks, to build robust
system and organizational solutions. Design thinking tools deliver strong
results in the stakeholder engagement process to connect user knowledge
and needs to the goals of this project. UPD’s use of this process during stake-
holder engagement sessions required users to think of the end state design
rather than the solution to a current problem or constraint.

Thus, we took on the task of identifying the key questions stakeholders
want to answer with data with the goal of making teacher talent data more
useful and powerful.

**The Chicago Convening, July 2016**

A group of stakeholders from TPPs, states, districts, and foundations were
invited to participate in a two-day convening in Chicago and as an ongo-
ing member of the team to provide feedback on use cases and the common
data standard. There were 35 participants across 19 states and 15 TPPs.
Where possible, “trios” of a TPP, SEA, and LEA were invited to partici-
pate together in order to leverage and encourage discussion around teacher
preparation data and data exchange in agencies that have been actively working together. In addition, individuals from institutions and organizations that have been forward thinking in using teacher preparation data for decision making were identified to participate in the discussions.

UPD Consulting designed and facilitated the stakeholder engagement sessions in collaboration with a team of teacher preparation subject matter experts. Session participants were grouped by into three functional areas based on the participant’s area of expertise. The groupings covered the three main perspectives of a TPP data user and were used as break-out groups to allow for richer discussions. Participants did not have to perform their grouping role as part of their daily jobs but instead were asked to assume this perspective for the purposes of the work. Each group is described below.

- Clinical: perspective of faculty or staff who coach teachers or teacher candidates, including overseeing student teaching (e.g., site coordinator, student-teacher supervisor, instructional coach)
- Coursework and curriculum: perspective of faculty in the education department who design content methods and curriculum to appropriately prepare teacher candidates to become quality teachers
- Administration: perspective of administrators who oversee functions that support education of teacher candidates (e.g., human resources director, talent manager, dean, teacher evaluator)

The process focused first on the decisions needed to be made by the TPP and then aligned the data elements to those decisions. This enabled UPD and the participants to develop clear use cases and ensure that the data standard will capture essential data elements most meaningful to the continuous improvement work of the TPP.

The use cases, which will be used to inform the definition of a TPP data standard, were created with the participants brainstorming the common questions TPP partners want to answer, using data from the perspective of their assigned role. UPD facilitators vetted each question with participants, working through the following process to increase understanding:

- Clarify what the decision is, when it is made, how often it is made;
- Clarify who typically makes the decision and who is a part of the discussion and decision;
- Clarify how data are used to make the decision and what the typical outcome is; and
- Understand how this decision supports continuous improvement in a TPP.

The participants also voted to determine the most critical decisions that helped establish priorities for the use cases and drove the discussion areas of focus for the remaining conversation. For the highest priority items (based
on voting), the facilitators identified the specific data elements that would support answering each identified decision. Specifically, the facilitators vetted the following:

- What are the data that you need to answer this question and make this decision?
- Where are the data commonly stored in terms of a system (e.g., SIS, online observation rubric software, etc.) and entity (LEA, SEA, etc.)?
- What are some of the common barriers to collecting, accessing, or using (analyzing, disseminating, making decisions on) the data?

Participants commented that the work described above was thought provoking and difficult, and most participants believed that work was setting the collective TPP community on the right path to a more comprehensive understanding of all the decisions and data elements essential and common to the field.

**The Use Cases**

During the convening in Chicago, participants identified more than 250 questions that they would want to answer with data. These data elements and questions acted as a starting point for the creation of the 91 use cases that would eventually serve as the foundation for building the TPDM. The following sections describe the process of honing the Chicago session information into a concise set of use cases, data elements, and data systems and sources list.

After the session, UPD captured all the data elements, questions, categories, and prioritization indicators that were noted and worked to convert these items into use cases (see Table 6.1) using the model of:

- I am a—Identifies the role of the person or people that make the decision or ask the question documented in the use case.
- I need—Describes the high-level categories of data needed in order to make the decision or answer the question documented in the use case.

<table>
<thead>
<tr>
<th>I am a</th>
<th>I need</th>
<th>In order to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher educator</td>
<td>• GPA</td>
<td>Determine if candidates are prepared to pass certification exams</td>
</tr>
<tr>
<td>Dean</td>
<td>• Course grades</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Teacher candidate assessment data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Teacher candidate observation data</td>
<td></td>
</tr>
</tbody>
</table>
In order to—Describes the decision or question asked for the purposes of continuous improvement within the TPP.

UPD held several sessions with a team of subject matter experts and members of the funder team to clarify specific questions and gaps that arose while consolidating and refining the use cases. In addition, the UPD team reviewed a set of use cases developed by Arizona State University for the iTeachAZ dashboards to identify any additional use cases and data elements appropriate for the national data standard work.

After incorporating this expertise, UPD provided a draft of the use cases to the Chicago convening participants and held three online meetings with participants. These sessions maintained the groupings that were used in Chicago (clinical, coursework and curriculum, and administration) to continue the conversation that occurred during the small group sessions. In addition, UPD facilitators asked participants to provide feedback on the following:

- Do the use cases accurately reflect the conversations that took place in Chicago?
- Is the priority assigned to each use case the right level?
- Are there any significantly new use cases or data elements not reflected on this sheet that need to be added?

The full set of use cases are available at www.updconsulting.com/teacher-preparation-data-model/.

Creating the Data Elements List

The use cases are the narrative that informed what data might be included in the TPDM. During the Chicago convening, stakeholders identified and discussed the types and categories of data that they need (e.g., applicant demographics). Following the session, a team of UPD technical analysts translated this information into a list of data domains (categorizations) and data elements (e.g., race, gender, date of birth, etc.).

The list of data domains and data elements is a combination of the current Ed-Fi standard and desired extensions (domains and elements specific to TPPs). The objective of this effort was to define how and where to extend the Ed-Fi data standard to include additional data elements needed for teacher preparation while leveraging the existing K–12 Ed-Fi standard where possible.

Building the Data Model

Using the listing of data elements, UPD technical analysts performed a gap analysis to the Ed-Fi data standard to determine what must be developed if a data domain or connection does not exist. The focus in this task was
to identify where there are gaps in data elements, tables, and data relationships in the Ed-Fi model that must be added or changed. In order to maintain integrity of the standard and minimize potential disruptions to existing Ed-Fi licensees, we prioritized leveraging the data definitions, semantics, and data relationships of the existing Ed-Fi model as much as possible. Where we could not reference an existing feature of the model, we identified extensions needed and developed a technical approach for providing that extension.

UDP also utilized a set of Ed-Fi extensions for talent management developed by The New Teacher Project (TNTP, 2016). This data model primarily provided input on the domains of observations and surveys. This data model was very flexible to relate a lot of different dimensions of surveys to many different actors. Some are filled out by teacher candidates, some are by principals about their residency, and some are course evaluation surveys.

Using the Ed-Fi data model as the template, we knew we would need to add a few main capabilities. First, we needed to add all the data dimensions required to address TPP data needs relating to preservice (teacher candidate) data. For these preservice domains, we worked to model the nomenclature and structure of the “student” dimensions of the Ed-Fi data model. We knew we needed to associate different roles (e.g., teacher candidate becoming a teacher) to the same unique ID. We also anticipated a fair amount of work to incorporate post-secondary institutions into the K–12 focused hierarchy of organizations (for example, to enable these postsecondary organizations to be the targets of enrollment). We extended the data model to be capable of longitudinally linking the records of preservice teachers to their corresponding record as an in-service teacher in the classroom.

In order to ensure alignment and integrity to the Ed-Fi data standard, a data panel was formed to provide guidance and decisions on key areas of the data model design. The data panel included representatives from the Michael & Susan Dell Foundation, the Ed-Fi Alliance, and UPD.

Figure 6.2 is a visual representation of the extensions to the Ed-Fi data model for teacher preparation. The entire set of teacher preparation data elements includes both parts of Ed-Fi core and the TPDM extensions.

Data Systems and Owners

As we developed the TPDM, we also knew it was critical to understand the TPP technology market and the various potential data sources. As is true in K–12 Ed-Fi implementations, with teacher preparation there is overlap of system vendors and data types which differ from institution to institution. For example, some TPPs use a deep and broad array of functionality in Chalk and Wire™, while others use only its assessment features and use other systems to track student records. As part of the project, UPD analyzed two main areas in relation to data systems:
Figure 6.2 Ed-Fi data model with TPDM extensions
What are the most likely data system sources and data system owners for various areas of the TPDM (e.g., teacher candidate assessments, teacher certification)?

What are the most commonly used system vendors across TPPs?

**Data Owners**

Table 6.2 includes the organization that would typically “own” each of the data categories included. Note that there is some crossover, particularly across SEAs and LEAs. Organizations can best realize the true potential of the TPDM when they are able to “connect the dots” of these data sources through data sharing and exchange.

**Data Systems**

For the TPDM to scale, what we know from Ed-Fi K–12 implementations is that TPP organizations must see the value in having access to integrated data and improving their data use competencies. In parallel, education organizations must ensure that source system vendors see a financial upside to applying the TPDM data standard to their proprietary products. Currently the demand for this type of integration is maturing, but the supply of system vendors that are integrated with the TPDM is just getting started. The success of this project will hinge on how well system vendors leverage the standard and how well TPPs insist that the source system products they procure adhere to the standard.

<table>
<thead>
<tr>
<th>TPP</th>
<th>SEA</th>
<th>LEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>• TPP Recruitment/ Applicant</td>
<td>• Completers Certification</td>
<td>• Mentor Teacher Demographic Data</td>
</tr>
<tr>
<td>• TC Demographic</td>
<td>• Completers Employment</td>
<td>• Mentor Teacher Performance</td>
</tr>
<tr>
<td>• TC Key Assessments</td>
<td>• Completers Survey</td>
<td>• PK–12 Student Performance</td>
</tr>
<tr>
<td>• TC Performance Assessment/Observations</td>
<td>• PK–12 District/School Data</td>
<td>• PK–12 Student Growth</td>
</tr>
<tr>
<td>• TC Grades/Transcript</td>
<td>• PK–12 Student Demographics, ELL, SPED</td>
<td>• PK–12 Student Perception</td>
</tr>
<tr>
<td>• TC Fieldwork Placement</td>
<td></td>
<td>• PK–12 Teacher Exit Data</td>
</tr>
<tr>
<td>• TPP Course</td>
<td></td>
<td>• PK–12 Professional Development</td>
</tr>
<tr>
<td>• TPP Faculty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Completers Demographic</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N.B.: TC = Teacher Candidate
In December 2017, the TPDM and supporting technical documentation was made available on the Ed-Fi Exchange (https://exchange.ed-fi.org/). This enables access to the TPDM technical artifacts by Ed-Fi community member organizations. The Ed-Fi TPDM can be used by TPPs, SEAs, and LEAs.

Figure 6.3 is a high-level representation of the structure of the TPDM. As part of an implementation, an organization may utilize the parts of the data model for which they have data available.

The TPDM is being maintained by the Ed-Fi Alliance as a “sponsored standard.” This means that the Ed-Fi Alliance recognizes the TPDM as aligned with the Ed-Fi data standard. It is not currently part of the Ed-Fi core but may be in the future after additional field implementations. UPD is providing continued support to the Ed-Fi teacher preparation community to request feedback, identify necessary updates, and release new versions of the TPDM. As part of this support process, monthly webinars are being hosted.

Figure 6.3 Representation of the TPDM
to engage the Ed-Fi teacher preparation community to discuss implementations and best practices and to solicit feedback for revisions.

TPDM Early Adopters

With the technical artifacts complete, we were able to investigate what an implementation would really look like across TPP, LEA, and SEA partners. Up to this point, the TPDM was largely theoretical, based on use cases articulated by teacher preparation stakeholders but not yet based on real data within an organization. This was the opportunity to actually test the TPDM to see if it was aligned with both the needs of an education organization and the data that they maintain and need access to.

Three early adopter projects were defined to participate in a process to assess their systems, processes, and resources to determine how they could benefit from the TPDM and specifically what it would look like to implement the TPDM for their organizations. For each of the early adopter projects, we wanted to ensure that organizations that will contribute both preservice and in-service data were included in a partnership. This provided the opportunity to not only test out the technical aspects of the TPDM but also investigate how a partnership involving a TPP, LEAs, and a SEA can integrate data to provide information about the preparation of teachers and how programs are meeting the needs of their district and state partners.

Leading up to the selection of three early adopters, we had many discussions with potential early adopter TPPs, states, and districts and the project funders. There were decisions around the structure of each early adopter project along with actual selection of programs. The selected structure was that the TPP would be the focal point and lead for the implementation along with one or two of their main district and/or state partners. The following organizations agreed to participate in the TPDM early adopter projects:

1. Texas Tech University with partners Dallas Independent School District and Lubbock Independent School District;
2. Relay Graduate School of Education with partners New Jersey Department of Education and New York City Department of Education; and
3. New Visions for Public Schools with partner New York City Department of Education.

Each TPP set the direction for the overall requirements and defined their specific needs for the TPDM within their organization. The process began with a data diagnostic that included the analysis of the organization’s technology architecture (i.e., the systems and data they maintain) and defining the data solutions necessary to implement the TPDM. UPD conducted interviews and focus groups with TPP staff (e.g., deans, recruiters, site coordinators, researchers) and technical leads to assess the systems, processes, and resources in relation to the TPDM. Analysis with the LEAs and SEAs
Ana Quintana, Bryan Richardson, Cari Reddick

focused on the data that they have available to share and contribute to the TPDM and the technical methods for data transfer. Through the data diagnostic process, UPD included recommendations for the most appropriate organization to host and maintain the Ed-Fi TPDM ODS.

Each of these projects was a three- to four-month engagement that included the following activities:

- Data diagnostic to review data systems, usage, and processes related to the TPDM;
- Gap analysis to identify critical gaps in the current state that would impact an implementation of the TPDM;
- Solution roadmap which includes recommendations, technical requirements, and an implementation plan including a project plan and budget to implement the TPDM; and
- Onsite review of recommendations and implementation plan to determine next steps.

Key Considerations for an Implementation

Within each of the early adopter projects, UPD and each education organization gained many insights about the data, systems, and resources and all the benefits and knowledge that could be gained by integrating data within their own organization along with strengthening the partnership across the TPP, state, and districts. Coming out of the three early adopter scoping projects, the major lessons learned were in the following areas:

- Data access—This includes defining the common key data that TPPs need access to in order to answer key questions; challenges that TPPs, states, and districts face in sharing and accessing this top priority data; variation across states on data access processes and policies; and best practices that have been established among some TPPs, states, and districts on sharing and reviewing key data for program improvement.
- Costs for implementation—There was much analysis and discussion around the costs and resources needed for implementation of the TPDM, as this is key to the sustainability and growth of the TPDM. While Ed-Fi is an open-source and free technology to use, there are costs associated with implementing the technology. For the first implementations, there are many risks in working with a new set of system vendors, the new data model, and new data sharing structures. UPD mapped out various scenarios on how to drive the risks and their associated costs down over time, including leveraging resources across implementations in the same state and with the same vendor system and use of the Ed-Fi application programming interface (API). The Ed-Fi API enables automating the data exchange between a source system and the Ed-Fi operational data store (ODS). To make this possible, a source system vendor maps
their database fields to the Ed-Fi TPDM and integrates with the Ed-Fi API. This creates efficiencies and cost savings across the field, because after this is completed, any other TPP that uses that system vendor does not have to manually map their source data to the Ed-Fi TPDM. The resource investment will continue to be monitored and tracked through the initial implementations, and lessons learned will be shared back to the Ed-Fi teacher preparation community.

• Organization best suited to implement the TPDM—Through the selection of the early adopters and through the scoping projects, there have been many insights about the characteristics of an institution or set of institutions that would benefit most from an implementation of the TPDM and the resources these organizations need to implement the TPDM. These learnings continue to be applied and refined through additional investigations with TPPs, coalitions, and state agencies that explore a potential TPDM implementation.

**Current TPDM Implementations**

Coming out of the scoping projects, there are four projects that are moving forward with TPDM implementations that represent different use cases of the model:

1. **State implementation:** New Mexico Department of Education was the first official implementation of the TPDM. The state utilized the TPDM to develop reports for TPPs across the state.

2. **Coalition of TPPs:** The University-School Partnership for the Renewal of Educator Preparator (US Prep) is a coalition of seven universities implementing the TPDM and dashboards for comparative analysis of their common measures. Texas Tech University is one of the partner universities, and after participating in the early adopter project, it determined that there was greater value in an implementation across the coalition of US Prep than for its specific program.

3. **TPP with several campuses in different states:** Relay Graduate School of Education is implementing the TPDM and dashboards across its 15 campuses. This project has started with a specific focus on data governance and data sharing to ensure they have the processes and data in line for a successful implementation of the TPDM.

4. **Individual TPP:** University of Texas Rio Grande Valley decided to pursue a data diagnostic scoping project on its own and is moving forward with an implementation.

The collective goals of these projects are:

• Build alignment across data systems and ensure that those systems are robust and connected through a central data store that contains data
from TPP, LEAs, and/or SEAs in the regions where teacher candidates and program graduates teach.

- Obtain data from LEAs and SEAs and link TPP data to construct a comprehensive picture of the TPP’s impact.
- Create dashboards and reporting structures that allow faculty and staff to frequently access data to inform decision making.
- Build meaningful opportunities for regular data review within and across campuses and construct learning communities across campuses that regularly examine common strengths and areas of growth based on data in an effort to learn and continuously improve.
- Provide technical, on-the-ground assistance in supporting TPPs in achieving a culture of habituated and systemic use of data to make decisions about teacher candidate interventions and program revisions and to ensure implementation fidelity and rigor.
- Share data in ways that allow stakeholders to utilize findings and learn from one another.
- Utilize data to inform technical assistance and support to individual providers.
- Conduct research investigating how the implementation of the TPDM and associated dashboards drives continuous improvement efforts.

There are also many other organizations that have expressed interested in pursuing an implementation as either an individual TPP, a coalition of TPPs, or at the state level. Our hope is that the use of the TPDM continues to grow and expand in order to realize the true value and vision of the project: to make the access and integration of data critical to TPPs easier and less expensive and to make innovative solutions developed in one location scalable to others.

The Next Phase: Collaborative TPP Dashboard

The Ed-Fi TPDM addresses part of the challenges in data access and use by providing a location to integrate data from disparate data sources. For this data to be used to guide program-level decisions, a method and tool to review and consume the data is necessary.

After developing the TPDM, UPD proposed that a similar collaborative approach be utilized to develop a dashboard and reporting tool that can be implemented by TPPs and state and/or district partners that integrates with the TPDM. This dashboard is being developed through a process of gathering requirements, feedback, and best practices from TPPs and partners. This dashboard tool will also be made available through the Ed-Fi Exchange (https://exchange.ed-fi.org/) along with the TPDM so that organizations may choose to implement out of the box or customize to tailor it to their needs.
The principal result of the Collaborative TPP Dashboard Project will be to provide TPP stakeholders with data they need to effectively and efficiently analyze how they are preparing candidates to become successful teachers that have a positive impact on student achievement and to help them make data-based adjustments to their programs to better prepare their teacher candidates for the students they will be teaching.

Why Collaborative?

Gathering requirements for and development of a dashboard and reporting tool is a time- and resource-intensive process, and in many cases, it is cost prohibitive for an individual institution to do on its own. Some programs do have existing dashboards or reporting tools, but they may not be meeting stakeholder needs because they are not tailored to specific use cases and do not provide a comprehensive view of all relevant data. Using a collaborative process to design this tool will result in cost savings as compared to each TPP doing this work in isolation. The TPP dashboard will provide a solution that has followed an extensive requirements-gathering, design, and development process, incorporating best practices in data review and dashboard design from across the teacher preparation field. This will not only serve programs that are involved in this project but also any institutions that plan to implement the Ed-Fi TPDM in the future, as these are provided at no cost to use through the Ed-Fi exchange.

Based on work in developing the TPDM, we know that there are many common data needs across teacher preparation stakeholders. For example, there are standard questions that a dean at any TPP would need to answer about their program and program enrollees. Therefore, we seek to develop this tool to address those most common, high-priority questions that stakeholders would use a dashboard to answer. We will also be developing the tool as highly flexible so that when organizations actually implement the dashboard, they can customize it to meet their needs. This includes branding, color scheme, and terminology that will ensure the tool “fits” with the organization and will aid in user adoption. Organizations will also be able to change or customize metric calculations and add additional visualizations or modules.

Inputs to Requirements

Coming out of developing the TPDM and early adopter scoping projects, we had a great base of knowledge of what data teacher preparation stakeholders need to address questions and make decisions for program improvement. The following are being used to provide input to the requirements for the TPP dashboard.
Use Cases

The use cases that were gathered to develop the TPDM are being utilized as a basis to define what questions the dashboard should be able to answer. Some of these use cases include:

- Assess the diversity and general demographic profile of currently enrolled teacher candidates;
- Assess if partner schools and students in those schools hosting teacher candidates are benefiting from partnerships;
- Identify areas where teacher candidates need interventions and remediation;
- Determine if program graduates are working in high-need districts or content areas post-enrollment;
- Evaluate how program graduates are progressing relative to other novice teachers;
- Determine how program graduates are demonstrating impact in the classroom based on student performance; and
- Review program graduate longevity in teaching by program.

Based on Best Practice

As part of the UPD investigation and as part of the first stakeholder convening, we reviewed and analyzed various examples of dashboards. This included state-level teacher preparation dashboards and reports (including Illinois, Louisiana, New Mexico, North Carolina, and Tennessee), individual TPP dashboards (including Arizona State University MLFTC, New Visions for Public Schools, Relay GSE, Teacher Squared, and Texas Tech University), K–12 district-level and state-level dashboards (including Shelby County Schools and Virginia), as well as examples from outside of education.

Stakeholder Engagement

Following a similar process as with development of the TPDM, we are engaging a group of teacher preparation stakeholders from many organizations and various roles within those organizations (e.g., deans, site coordinators, faculty, teacher candidates, researchers, recruiters) to provide feedback on dashboard requirements, including data views, measure definitions, and user access levels. Many of the same individuals have been engaged in this process since the beginning of this journey back in Chicago in 2016. In addition, many individuals from US Prep, Relay GSE, and University of Texas Rio Grande Valley (UTRGV) who are currently working on their TPDM implementation project are involved. These stakeholders have the added perspective of knowing they will be using this tool and envisioning how they will use the dashboard on a regular basis.
As of May 2018, we have held two in-person convenings and conducted many virtual sessions with stakeholders. During the first convening in November 2017, we gathered input and expertise on how, when, and where stakeholders use data. Participants provided feedback on what is working well and what needs improvement for any existing dashboards and reports that they use, what they would like to see in a dashboard, who needs to use the dashboard, and how user access should be managed. During the session, groups of participants sketched out initial wireframes for the dashboard for the various user roles.

In preparation for the second convening, the UPD team is developing a site map and series of wireframes with sample visualizations across all modules of the dashboard. The site map is organized by the following modules: recruitment to enrollment, candidate performance, program information, post-completion, and university-school partnerships. During the session, participants reviewed these wireframes to address key questions: What is the highest priority content to include on the dashboard? Is this the best way to visualize the data? Is the data actionable? In addition, we reviewed sample metrics to gather input on special considerations for defining calculations.

Through an iterative process, UPD will continue to review and develop the dashboard tool with stakeholders for further refinement.

**Dashboard Design**

The TPP dashboard is being developed to integrate with the TPDM (see Figure 6.4). This means that each of the metric calculations are written using standardized data elements from the TPDM. As part of the process of implementing the TPDM, an organization goes through a data mapping process in which they map their source system data to the standardized TPDM. Once this data mapping is done, the organization’s data is standardized and the TPP dashboard can be implemented with the TPDM.

![Figure 6.4 TPP dashboard connected to Ed-Fi TPDM ODS](image)
Next Steps

As of May 2018, the TPP dashboard project is still under development. After completing an internal and external testing process, the dashboard tool and technical documentation will be available on the Ed-Fi exchange so that it can be utilized by all members of the Ed-Fi community.

Once the dashboard is implemented with the early adopters, we will have the ability to demonstrate the power and benefits of the TPDM and dashboards, a concept that has been abstract and difficult for some to conceptualize. The dashboard, along with the Ed-Fi TPDM, provides a solution to bring together both data from disparate systems and the tools to review that data for the purposes of driving action for program improvement. Providing the technical tools to support data use by more programs will enable them to improve support for individual teacher candidates and address overall program improvement.

Looking to the Future

During the March 2018 dashboard convening, Michelle Franco-Westacott (personal communication, 2018), Regional Transformation Specialist with US Prep, commented at the end of the session that it is so exciting to see how this work has evolved over the last two years. In July 2016, stakeholders were creating inventories of questions and data elements on flip charts and Post-it Notes, and we had many discussions to just establish a common language and vocabulary across this data. Now, in May 2018, a physical data model exists in the form of the TPDM that is available to any teacher preparation program, there are multiple implementations in progress, and we are turning the data into usable and meaningful information for stakeholders through the TPP dashboard. During this same convening debrief, Sarah Kolbe (personal communication, 2018), Educator Preparation Data Scientist at California State University, commented about “Phase 3,” where we can really start to understand what programs are learning by having access to this data and how the data is being used in the field for program improvement.

The key success factor of realizing the benefits of the investment in this work is the pathway to scale across the teacher preparation space. In practice, the TPDM can be implemented in many ways: for a coalition of TPPs, for an individual TPP with their state and district partners, or across a state including the SEA and TPPs. The benefits of having a standardized data model is apparent when many organizations are utilizing the TPDM and dashboards, improving upon these tools, and sharing back advancements and best practices to the teacher preparation community.

The Bill & Melinda Gates Foundation, the Michael & Susan Dell Foundation, the Ed-Fi Alliance, US Prep, Relay, UTRGV, and UPD are working to build a culture of consistency, awareness, and sharing practices across a
community in order to drive the cost down, bring vendors to align their systems to the TPDM standard, and encouraging states and districts to share data, with the final goal of increasing the effectiveness, adoption, and sustainability of the solution for the teacher preparation community.

References


Overview

In 2011, Relay Graduate School of Education (Relay GSE) was founded “to teach teachers and school leaders to develop in all students the academic skills and strength of character needed to succeed in college and life” (Relay GSE, 2018). Relay GSE is a fully accredited, nonprofit, independent institution of higher education that now teaches students in 16 campuses across the nation. When Relay GSE first began, one of the questions its founders pondered was the following: “If a teacher is teaching, but students aren’t learning, then is that teacher really teaching?” It was in this spirit that Relay GSE was founded as the first graduate school of education in the nation requiring its graduate students—all of whom were working as full-time teachers—not only to demonstrate proficient teaching capacity via their coursework but also to prove that their own PK–12 students were learning. The brainchild of Brent Maddin, Relay’s director of teaching and learning (later to become the school’s first provost), Relay GSE stipulated that, in order to earn a master’s degree, all graduate students would have to provide evidence that the PK–12 students they were teaching had achieved a meaningful amount of learning—approximating one year’s academic learning in one year’s time—as a result of their instruction. Relay GSE asked graduate students both to collect data about how much their students were learning throughout the year and to use those data for continuous improvement. The school also asked graduate students to consider what these data indicated they could improve in their own teaching practice as well as what these data told them about individual students who might need more—or different—supports in order to succeed at high levels.

Though we were excited about the pioneering aspect of our work, the process of creating, implementing, and supporting Relay’s SGA curriculum was not without obstacles and pitfalls. After our first year rolling out the new Student Growth and Achievement (SGA) curriculum, graduate students rated this curriculum as one of the most time-consuming and unpopular aspects of the entire Relay GSE program. Faculty interviews and water-cooler conversations confirmed that our faculty members were having an equally
challenging time teaching and supporting SGA content. Thankfully, we saw these data as a starting point for improvement. If we were asking our graduate students to be data-driven in their own work, then certainly we needed to hold ourselves to that same standard. This chapter is the story of how we applied the principles of continuous improvement in order to transform our SGA content from being a big headache for students and faculty into being content that over 70% of faculty said they either liked or loved to teach and that alumni consistently rated as being Relay GSE’s biggest value-add. This is a story of learning to walk—not just talk—the knowledge, skills, and mind-sets of using data for continuous improvement.

Background

Historically, graduate schools of education have conferred degrees upon teachers based on inputs more so than on outcomes (Levine, 2006). Additionally, there is little evidence to suggest that the mere possession of a master’s degree improves teacher skill (Hanushek, Kain, & Rivkin, 2005). To address these fundamental issues, Relay GSE has asked its teachers to demonstrate their ability to lead students to academic achievement (SGA) as a graduation requirement (Honigsberg & Salmacia, 2016). In order to graduate with a master’s degree from Relay GSE, graduate students need not only to demonstrate proficiency on their course assignments but also to lead their PK–12 students to demonstrable academic gains over the course of the school year. In order to support graduate students in achieving this ambitious aim, Relay GSE created the SGA element, which became the centerpiece of Relay’s coursework (see Figure 7.1).

Through SGA coursework, graduate students were given the opportunity to develop a deep understanding both of how to measure student learning within their teaching contexts and of how to help their own students achieve their academic goals. Graduate students were taught a five-step pathway for measuring student achievement that was both general enough for any teacher to use and specific enough to be a high-integrity process for all (see Figure 7.2).

SGA coursework followed a five-step pathway for measuring student achievement, whereby graduate students:

1. determined the content their students needed to master by end-of-year;
2. solidified an assessment plan to gather evidence of mastery of that content;
3. set quantifiable goals for student learning;
4. tracked and analyzed student progress toward those goals; and
5. verified student outcomes at the end-of-year.

The SGA element comprised two years of coursework leading up to an oral defense of their PK–12 students’ academic and character gains. In the first
year of SGA coursework, graduate students practiced applying the five-step pathway in a low-stakes environment; in the second year of the program, graduate students were responsible for measuring academic and character outcomes for the students they were teaching that year.

Making SGA a critical component of earning a master’s degree in education highlights the founding vision of Relay GSE and was a revolutionary change in the field of teacher education (Salmacia & Honigsberg, 2016). To oversee the creation and rollout of this curriculum, a team of two was created, led by the authors. 1 We became known as “Team SGA” and were charged with making good on Relay’s larger vision. The SGA curriculum represented our attempt to take that abstract vision and transform it into an effective, sustainable, and scalable process. This revolution, however, was not without its challenges. We quickly learned that the data-literacy content within SGA coursework was challenging both for our faculty to teach and for our graduate students to learn. Our initial data trackers, which graduate students used to store and explore their student achievement data, were difficult to use. This difficulty caused frustration among graduate students and
The RGSE Pathway for Measuring Academic Achievement

Step 1: Determine Content

Select Subject, Students, and Measurement Approach

Qualifying Qs:
1. What are you measuring and for whom?
2. Are you using a mastery- or growth-based approach?
3. Have you selected an aligned student achievement tracker?

Step 2: Solidify Assessment Plan

Create Plan to Assess Mastery of the Content

Qualifying Qs:
1. Have you selected the right assessment(s)?
2. When do you plan to administer your assessment(s)?

Create Plan to Measure Growth in the Subject

Qualifying Qs:
1. Have you set meaningful goals for your class?
2. Have you set meaningful goals for all students?

Step 3: Set Goals

Confirm Mastery-Based Goals for Student Learning

Qualifying Qs:
1. Have you set meaningful goals for your class?
2. Have you set meaningful goals for all students?

Confirm Growth-Based Goals for Student Learning

Qualifying Qs:
1. Are you maintaining complete, error-free data?
2. Have you analyzed your data?
3. Have you responded to your data?

Step 4: Track Progress

Collect and Act On Ongoing Mastery Data

Collect and Act On Ongoing Growth Data

Step 5: Verify Outcomes

Complete and Reflect Upon Mastery Data

Complete and Reflect Upon Growth Data

Figure 7.2 Relay GSE pathway for measuring academic achievement

took time away from other important teaching duties. Furthermore, the entire SGA process was incredibly meticulous and time-consuming, requiring a large amount of effort to pull off relative to the other coursework faculty members were teaching. In short, we had what we thought were good ideas for SGA curriculum, but we were struggling to implement those ideas with fidelity and ease.

When we began to collect data about the struggles our faculty and graduate students were facing with our SGA curriculum, we knew we needed to act, and act fast. We knew things needed to get better, but we weren’t entirely sure what that meant or how to do it. We both agreed, however, on the need for a systematic and deliberate approach to improvement. Then we thought, “Why not apply the same data for continuous-improvement principles that we ask our graduate students to use in their SGA coursework to our own work? What if we used data about our own work as the basis for our own continual improvement?” While we knew our SGA conceptual framework wasn’t the right approach for our work with faculty, we suspected that we could draw upon continuous improvement frameworks from the business world to guide our work. It thrilled us to discover that, whereas we’d had to design our SGA curriculum from scratch, there was already an existing framework for a continuous-improvement process. Next, we set out to learn about that framework and put it to use.

The DMAIC Framework for Continuous Improvement

If you already drive a Toyota, there is a nearly 70% chance that your next car will also be a Toyota. Given this impressive degree of brand loyalty, highest among all multi-brand automobile manufacturers (Miller & Smith, 2017), it is difficult to believe that there was ever a time when Toyota had to look to an American car manufacturer for ideas to improve its own manufacturing process. However, the Toyota Company was struggling with production efficiency and realized it could be helpful to look outside of its own organization to find an established solution to their problem. Looking outside of its own organization is exactly what Toyota did. Toyota learned from Ford Motor Company, the leader at that time in rate and volume of automobile manufacturing, and incorporated elements of the Ford process in order to improve Toyota’s quality and output (Baghel & Bhuiyan, 2005). As we approached the work of improving the SGA experience for faculty and graduate students, we took a page out of Toyota’s book and looked to the broader field for approaches to improvement that we could use.

That search led us to learn more about the six sigma and the DMAIC model used to implement the six sigma philosophy. Early adopters of cell phones might remember the rapid growth of the Motorola Company and
Learning to Walk the Walk

Motorola is a company credited as a pioneer in six sigma philosophy, a philosophy that is reported to be responsible for achieving huge cost savings while supporting growth (Baghel & Bhuiyan, 2005). Six sigma is data-driven philosophy that “reduces waste, increases customer satisfaction, and improves processes” (Adams, Jones & Paras, 2010, p. 416). Six sigma is implemented through a process called DMAIC: define, measure, analyze, improve, and control. DMAIC “is widely used when a product or process is already in existence but performing inadequately,” (Adams et al., 2010, p. 416). When we found the DMAIC model we knew we’d found a good fit for our needs, as we already had an SGA process in place but needed to improve the quality of that process. The DMAIC framework grounded us in a continuous improvement process that was deliberate, methodical, and appropriately staged. For us, the spirit of the DMAIC approach encompassed everything we were hoping to do: define our problems, measure where we currently stand and what and how much we want to improve, analyze the causes behind our problems, improve upon our current state by creating and testing new solutions to our problems, and control the gains we achieved once we had achieved them (Basu, 2004).

Although the DMAIC model was created outside of an education context, we found that the model provided a set of organizing principles that were exactly what we needed. The DMAIC model charts a sequential pathway for improvements to a process, and it articulates the scope of the improvement cycle. Most importantly, the DMAIC model was grounded in a mind-set in which we believed deeply—that data help us improve. In the sections that follow, we outline the steps we took in relation to each of the DMAIC phases in order to address the challenges we faced in rolling out our SGA coursework. We begin by briefly describing the purpose of each phase and follow with a detailed description what we did in response.

**Define**

The first stage in the DMAIC approach is to define to the problem at hand and to determine the scope of the improvement project (Basu, 2004). Continuous improvement cannot commence effectively unless stakeholders across the organization agree on what they are improving, on the nature of that improvement, and on how much improvement they seek to generate. This is also a prioritization and investment stage, wherein the organization must decide how important a problem really is and must confirm that they are willing to devote the resources to the subsequent improvement effort (Shankar, 2009).

To better understand our problem, we engaged in a broad exploration of what graduate students and faculty members were feeling and experiencing...
in SGA coursework. We reviewed graduate student and faculty perception survey data, met with faculty members, and held focus groups with graduate students. We also observed how SGA coursework was playing out in real classrooms by observing SGA class nights and by guest-teaching SGA courses. We triangulated the quantitative data we received via our surveys with qualitative survey responses, interview takeaways, and our observations from the field in order to understand the problem in greater detail. At the conclusion of this investigation, we identified several big problems we needed to solve:

- Many graduate students did not like taking SGA courses;
- Many faculty members did not like teaching SGA courses;
- Data trackers used to store and explore PK–12 student achievement data were cumbersome and broke easily;
- Graduate students and faculty members were overwhelmed by the perceived complexities of the SGA curriculum;
- There was variation in how the SGA curriculum was being taught by different professors;
- SGA course assessments were scored inconsistently; and
- Overall, SGA coursework felt too time consuming for graduate students and faculty members.

After identifying our biggest problems with the SGA coursework, we identified that the SGA team was the group of people who should work to solve these issues with support from the provost and dean. We decided on a time line of one school year to stage our improvement efforts and measure their results. At this point, we identified problems that were mostly about negative perceptions surrounding SGA coursework and inconsistent delivery of SGA content. We felt a need to act quickly and received advice about what we should do—suggestions such as selecting a designated professor to teach all SGA courses, or moving SGA coursework to our content Saturday classes to be taught by our part-time PK–12 content faculty. This sort of immediate jump to action, however, is the exact misstep that the DMAIC model seeks to prevent. We had some ideas about what we wanted to do next, but we knew we needed to formalize our improvement goals, analyze our data, and determine our root causes before we could begin creating solutions.

**Measure**

The next step in the DMAIC approach is to determine what data to use to measure improvement efforts and to set quantitative goals. The measure stage is when baseline performance metrics are established to describe the current state of the program (Basu, 2004). While the define stage helps to
articulate generalized problems with the system, the measure stage seeks to quantify those problems with concrete data. These measures serve as a North Star for the continuous improvement efforts. The measure stage is also when the stated aims of the define stage can be revised to reflect truly measurable outcomes. After better defining our problem, we knew we wanted to set measurable goals relative to three key areas:

- Improving graduate students’ experience with data trackers;
- Increasing faculty knowledge and comfort with teaching SGA courses; and
- Increasing consistency of SGA instruction and scoring of SGA course assessments.

One challenge we encountered at this stage was deciding what performance measures would reflect a successful improvement. We had graduate student perception survey data about single SGA sessions they had attended, as well as anecdotal evidence that data trackers were incredibly challenging to use, but no other quantitative baseline data from which to set improvement goals. We realized we lacked the necessary data-collection mechanisms to really measure improvement. In our reading, we learned that we were not alone in this struggle. Continuous improvement efforts are often hamstrung by the limitations of “quality data, measurement, and reporting” (Antony, Douglas & Fryer, 2007, p. 508). Rather than letting this lack of data systems get in our way, we decided to move forward with what data we did have, created new data-collection processes, and set ambitious performance goals for data we had yet to collect. We ended up setting the following quantitative goals for performance that we intended to achieve in one year’s time:

**Improving Graduate Students’ Experience with Data Trackers**

- 80% of graduate students Agree or Strongly Agree that they know how to use their data trackers
- 80% of graduate students Agree or Strongly Agree that they received the support they needed with their data trackers

**Increasing Faculty Knowledge and Comfort with Teaching SGA Courses**

- 90% of faculty members will Agree or Strongly Agree that the SGA team provided them with high-quality training
- 90% of faculty members will Agree or Strongly Agree that the SGA team provided them with high-quality support
Increasing Consistency of SGA Instruction and Scoring of SGA Course Assessments

- 100% of faculty members will demonstrate proficiency of SGA measurement principles before teaching SGA coursework
- SGA team will agree with faculty decisions 80% of the time during SGA course-assessment audits

As before, it was tempting to move directly from goal-setting into problem solving. At the conclusion of the define phase, we knew that we wanted to solve problems related to the negative perceptions about SGA coursework and the inconsistent delivery of SGA courses. At the conclusion of the measure phase, we also had a way of quantifying and operationalizing those ideas with the use of survey data and quality-assurance metrics such as course-assignment audits. What we lacked still was the “why” behind those problems, which would unveil the “how” when it came to solving them. Why were these survey data so unfavorable? Why were faculty members frustrated with teaching SGA content? These questions led us to further analyze the causes of our problems.

**Analyze**

In the DMAIC approach, the analyze stage represents a methodical investigation of root causes, presenting an opportunity to hypothesize and then pressure-test beliefs about why the problems exist (Basu, 2004). Analysis of a problem should attain sufficient depth and detail to uncover variation in the problem and to identify causes of that variation. For example, even if the define phase identifies that overall customer satisfaction is low, there are still usually a few satisfied customers within that aggregate, or at least a few who are satisfied with some aspect of their customer experience. The analyze phase pays careful attention to both the experience of the majority and any differing trends for the minority, especially positive deviations. “Why” questions and fishbone diagrams are great tools for effective analysis (Bryk, Gomez, Grunow, & LeMahieu, 2015).

We began our analysis by digging into the qualitative data we collected about the problems surrounding our SGA coursework. Using qualitative data from interviews, focus groups, observations, and guest teaching, we sought to parse out the causes of the specific problems our graduate students and faculty members were having with SGA coursework. We also carefully consulted with the graduate students and faculty who held very positive perspectives about the SGA curriculum in order to understand what caused them to have a favorable experience. A closer analysis of the data revealed the following prioritized causes of our problems:

**Graduate Students and Faculty Felt “frustrated”**

- Graduate students were having to spend a large amount of time entering student achievement data into Relay GSE trackers, and these trackers
were frequently breaking down and needed to be sent to faculty to get fixed.

- Faculty members didn’t know how to fix the student achievement trackers and were frustrated that they couldn’t better support their graduate students.
- Graduate students had many questions about the SGA process (e.g., which standards to select, what assessments to use, what to do when a new student joined their class midyear) that faculty felt underprepared to answer, and there was no clear place for faculty members to go for help.

Faculty Felt “stressed” and “anxious”

- Many of our faculty had not measured student outcomes in their own tenure as PK–12 teachers and expressed a lack of comfort with the idea of teaching others how to do something they had never done.
- Some faculty members expressed a lack of familiarity with the data-literacy skills we were teaching our PK–12 students.
- Many faculty members had not used Microsoft Excel in the past (the platform for our student achievement trackers) and shared mind-set and skill hurdles associated with using that software.

Faculty Felt “under-supported”

- We didn’t have a process or a tool for determining which faculty members felt knowledgeable and comfortable with teaching SGA courses and which didn’t.
- SGA course grading lacked sufficient detail and systemization to ensure accuracy and consistency across faculty members.

Our root cause analysis led us to believe that the root of our issues—and therefore the root of our solutions—lay with faculty knowledge of and comfort with teaching SGA content. We were not only excited to have come to a systematic conclusion that faculty knowledge and comfort was at the root cause of our problems, but we were also thrilled to figure out that our best solutions were likely to live within our faculty members. We have the pleasure of working with amazing and incredibly dedicated faculty members who were all exceptional teachers of PK–12 students and who brought that same level of skill and dedication to their training of novice teachers. We realized that we simply hadn’t done enough to arm our faculty with the knowledge and skills they needed to teach this new SGA content effectively and efficiently. We also knew this solution would be fun. Our faculty colleagues were a joy to work with, and we were thrilled about the opportunity to get them as excited about teaching SGA courses as we were.
Improve: Round One

Finally, we were ready to get to work on improving our stated problems and achieving our stated quantitative goals. The improve phase of the DMAIC model is an opportunity to problem solve while working with a concrete set of improvement initiatives that can be carried out with fidelity (Basu, 2004). Both planning and action occur in this phase. The planning work involves defining the scope of improvement projects in order to know what resources are required and establishing clear bounds for the project so that it does not continue indefinitely—“continuous improvement” does not imply that the improvement goes on forever! The improve phase is finite, as is the entire DMAIC process, so that the improvement initiatives will follow directly from the improvement plans.

Based on the root cause analysis we had completed, we embarked on the following improvement actions, for which we created project plans and completion timelines:

Improvement to Products

MADE SMALL FIXES TO DATA TRACKERS

The data trackers we asked our graduate students to use were not best in class, and we initially had some very lofty ambitions about how we might reinvent these tools from the ground up. If trackers were causing us big headaches, why not completely retool and improve them? In response, we surveyed the field, piloted a couple existing data-tracker tools, and conducted enough market research to realize that it was going to be very time consuming and very expensive to create the best data tracker we could or to configure any existing tool to make it optimal for our purposes. The resources and timeline required to meet our initial—and rather ambitious—idea proved to be untenable.

Once we better understood our constraints, we realized the actions we had the capacity to take involved making small fixes to our data trackers. Our tracker technology was intended to easily support—not readily overburden—the process of measuring student growth and achievement. Originally, we had constructed the tool with a full range of features intended to meet the needs of more sophisticated users. After seeing the trackers used in action, however, we realized the fewer bells and whistles, the better. Accordingly, we made small modifications to streamline tracker functionality for the everyday user. We took steps to eliminate the risk of error, reduced the number of permissible formats for inputting data, inserted drop-down options wherever possible, and were strategic with our use of data validation fields. Although these small fixes restricted some of the more advanced tracker functions, we hoped they would make the tool more intuitive and less cumbersome for the majority of users.
Improvement to Training

CREATED SGA FACULTY TRAINING FOR SGA

For a teacher, nothing beats the thrill of seeing students get excited about a topic they once considered boring. We believe one way to do this is to meet students at their zone of proximal development and used this approach to onboard our faculty members to the SGA content. We created a two-day training event for all new faculty and hosted this training in the summer, shortly after most new faculty had joined Relay GSE staff. During our training, we wanted faculty members to gain three things:

- knowledge and comfort in teaching SGA courses;
- knowledge and comfort in using and supporting graduate students’ data trackers; and
- a love of all things SGA.

Using three case study graduate students, we took faculty members through a simulated yearlong process of measuring their PK–12 students’ achievement, and we worked to ensure that faculty members had the necessary background knowledge and skills to respond to their graduate students’ questions with greater ease. We exposed faculty members to the SGA coursework they would be teaching, and we taught faculty members how to fix tricky tracker problems and respond to graduate students’ support requests. We worked to build faculty members’ confidence with the data tools, and we allowed them to feel success in their work via a gradual onramp of complexity. Our goal was for faculty members to leave the training with a sense of confidence and a belief that they could be their own graduates students’ “SGA and tracker heroes.”

We also knew that many faculty members didn’t look forward to teaching SGA courses and that developing knowledge and skill alone would not be enough to change this sentiment. To be successful in our change efforts, we knew we were also going to have to transform the culture surrounding the teaching of SGA courses. Through our review of higher-education change literature, we learned that

values become deeply instantiated in all human processes and structures. Thus, to change a method of teaching is not as simple as knowing the new mode of teaching one wants to put into practice; it also means unlearning the values associated with the existing mode of teaching.

(Kezar, 2014, p. 33)

Bolman and Deal (2013) also add that, “Our own cultural ways, ‘how we do things around here,’ are often invisible to us because we see them simply as the ways things are—and ought to be” (p. 244). These learnings let us
know that we had to make SGA something that faculty enjoyed teaching and that this was simply how folks should “feel it ought to be.” In addition to skill-building, we led intentional cultural change through the use of play, humor, initiation rituals, and the development of heroes. The new faculty training we held each year had its own playful theme that often included costumes, thematic games, and prizes. Faculty began to look forward to seeing what theme we selected for the following year, and we asked returning faculty to weigh in on our decision. We incorporated humor throughout our sessions—particularly in our tracker trainings, which we knew could be the place where faculty members were likely to get the most frustrated. We held a brief, team-based competition for new faculty trainees at the end of our two-day training and asked faculty to do silly things like stand on desks in our main office work space and proclaim that they loved trackers, or create and sing songs about the five-step SGA pathway. Over time, these silly activities became initiation rituals and rites of passage into the work of teaching SGA. Most of all, we did not take ourselves too seriously. We had fun, shared knowledge and resources, and trusted that, with these resources and supports, our faculty members would make the SGA content shine for their graduate students.

Improvement to Support

CREATED SGA AND TRACKER-SUPPORT QUEUES

Though it was our intention that all faculty members walk out of our two-day new faculty training armed with the knowledge and skills to be successful teachers of SGA content, we knew there were certain to be questions about SGA content and trackers that would come up during the year—ones that faculty would need support in answering. During our analysis stage, one of the causes of faculty frustration we discovered was that there was no clear place for faculty members to go to get rapid support when they had questions about SGA content. We prepared our faculty members to be quick responders to their graduate students’ questions, but we knew they needed their own places to go to for help as well.

In response, we created two email queues where faculty could turn for SGA support. The first was a place where faculty members could ask any questions they had about SGA content or curriculum; the second, a tracker queue where faculty and graduate students could go for help if they were having a problem with their trackers. The authors manned the tracker queue, and we both made a promise to respond to people thoroughly and within 24 hours. In Boston, they have an expression called “one touch,” which refers to the act of completing something in a single sitting—a particularly ravenous diner, for example, might “one touch” a pizza. We set a goal that we would “one touch” at least 90% of our help requests, answering them so thoroughly and clearly that no further clarification would be
needed, and that we would always do so within 24 hours. The ultimate vision for these queues was to build capacity in faculty members, while also acting as a Band-Aid to resolve existing problems for graduate students effectively and efficiently. We provided answers that both solved the immediate problem and provided information to identify any systematic issue that might surface again.

**CREATED A BIWEEKLY SGA DIGEST**

We also knew that it was important to keep in touch with faculty members after new-faculty training. Not everyone was going to write the queues regularly, and we wanted to have a way to stay in touch with faculty and keep them up to speed about all things SGA. We titled the newsletter the “SGA Digest”—in it, we included information about new instructional materials, reminders about upcoming courses and due dates, and we addressed questions graduate students might be likely to ask at a particular point in the year. We also took the SGA Digest as an opportunity to continue with our cultural change efforts and, therefore, made these digests lighthearted and fun. We used the digest to “shout-out” faculty members who had recently taught an exceptional SGA session or served as someone’s tracker hero, and we worked to highlight all the great efforts faculty members were making and the great results they were achieving. At the end of each SGA Digest, which went out approximately every other week, we included a tracker-fix challenge: identifying one of the trickiest tracker fixes that had come our way that week and offering a prize for the first faculty member who could solve the graduate student’s problem. We were constantly sending fun prizes to faculty members located in regions across the country, and we worked to keep faculty members engaged and excited about SGA content even when we didn’t have the chance to see them face to face on a regular basis.

**Improvement to Fidelity of Implementation**

**CREATED FACULTY SGA ASSESSMENT**

Our observations in the field and our root cause analysis also showed us that faculty members had varying levels of proficiency in the SGA content they were teaching, but we didn’t have a good way of knowing what particular content which individual faculty members were struggling with—and, therefore, did not know to whom to reach out in order to provide additional support. In response, we created a short SGA faculty assessment that simulated the knowledge and skills a faculty member would need in order to teach SGA coursework successfully. There were several purposes for the SGA faculty assessment, the most important being that we wanted to develop faculty members’ confidence with the content. We believed that if faculty members took and passed the SGA assessment, this would help instill
the self-efficacy needed to teach the content well. We wanted to give faculty a way to believe that they were ready for the task of teaching SGA beyond our own cheerleading of their work. We also needed a way to identify who still needed more coaching and support to be ready to teach SGA content so that we could provide them with additional supports before they began teaching graduate students. Despite our best intentions, we knew that a few folks would leave new-faculty training needing additional supports, and we wanted a way to quickly identify those people and get them the supports they needed. Sometimes we learned that a couple of faculty members needed some additional support with fixing trackers, while other faculty members needed additional support in measuring student growth and achievement for subjects they didn’t teach as a PK–12 teacher and were therefore less familiar with. It’s important to note that the SGA faculty assessment was used solely in a formative manner and in no way was used to evaluate faculty members. Whenever we identified that a faculty member needed extra support, it was a positive thing. We were able to follow up with the just right supports in a timely manner, making sure that we didn’t leave a single faculty member behind before the start of the school year.

BEGAN SGA COURSE AUDITING

Another important aspect of the SGA coursework was that we had exacting standards for the scoring of graduate student work. The reason for this was not to be more stringent than we needed to be but rather because graduate students were creating yearlong measures of their PK–12 students’ achievement, and errors that were present in those plans at the beginning of the year were certain to catch up with them by the end of it. In some cases, errors at the beginning of the year could even prevent a graduate student from successfully graduating on time. The stakes were high, and the need for meticulousness was pressing. During the define stage of our work, we discovered that faculty members were not scoring assessments consistently from cohort to cohort. As had been the case with our faculty-assessment scenario, we realized we did not yet have an efficient way to distinguish the people who were scoring course assessments accurately from the people who weren’t and therefore needed additional and/or different supports. We knew the first step to improving the problem of scoring consistency was to gather more information about the scope of the problem and to create solutions from there.

In response, we instituted an SGA assessment auditing process, in which we would randomly select a small percentage of graduate student course assessments to audit within each faculty member’s cohort. We scored the selected course assessments against the same rubric faculty members used to score the assessment, and then we compared our scores to the faculty members’ scores. Next, we emailed faculty members the results of their audit, which was a rundown of how our scores and feedback compared. If we
had different scores and/or glaringly different feedback for graduate students, we shared the reasons why we believed this was the case. All of our feedback once again was shared in a formative manner, and in no way did the process constitute an evaluation of any faculty member. Rather, we used this information to better inform the score-norming guidance we provided faculty members moving forward, hoping our formative feedback to each faculty member would help them score with greater accuracy in the future.

**Improve: Results of Round One**

After our first improvement cycle, we were eager to see the results of our efforts. We had received positive anecdotal evidence of improvement through word of mouth from our faculty friends, as well as some initial indication about our results from the faculty assessments we’d scored and the graduate student assessments we’d audited. When we documented progress toward our six improvement goals, we found that we’d met two of them (see Table 7.1).

While we were initially disappointed to see that we didn’t meet more (or all!) of our six goals, we found our first round of data to be encouraging. First, we had incredible success with our faculty training and support goals, which reflected improvement in our root cause area of focus. These data

**Table 7.1 Round One Progress toward Goals**

<table>
<thead>
<tr>
<th>Goal</th>
<th>Results</th>
<th>Met?</th>
</tr>
</thead>
<tbody>
<tr>
<td>80% of graduate students Agree or Strongly Agree that they know how use their data trackers</td>
<td>55% of graduate students Agreed or Strongly Agreed (n = 273)</td>
<td>No</td>
</tr>
<tr>
<td>80% of graduate students Agree or Strongly Agree that they received the support they needed with their data trackers</td>
<td>74% of graduate students Agreed or Strongly Agreed (n = 326)</td>
<td>No</td>
</tr>
<tr>
<td>90% of faculty members will Agree or Strongly Agree that the SGA team provided them with high-quality TRAINING</td>
<td>100% of faculty members Agreed or Strongly Agreed (n = 11)</td>
<td>Yes</td>
</tr>
<tr>
<td>90% of faculty members will Agree or Strongly Agree that the SGA team provided them with high-quality SUPPORT</td>
<td>91% of faculty members Agreed or Strongly Agreed (n = 11)</td>
<td>Yes</td>
</tr>
<tr>
<td>100% of faculty members will demonstrate proficiency with SGA measurement principles before teaching SGA coursework</td>
<td>77% of faculty demonstrated proficiency on SGA faculty assessment (n = 26)</td>
<td>No</td>
</tr>
<tr>
<td>SGA team will agree with faculty decisions 80% of time during SGA course-assessment audits</td>
<td>SGA team agreed with 70% of faculty scoring decisions (n = 50)</td>
<td>No</td>
</tr>
</tbody>
</table>
told us we were on the right track with regard to our improvement strategy. Our faculty now felt they were receiving high-quality support and training from the SGA team; next, we had to ensure that this training and support would translate into greater faculty proficiency with SGA content. Second, because we did not have baseline data from which to set growth goals, nor a clear understanding of what our baseline metrics would be for these goals, we erred on the side of setting ambitious quantitative targets. We wanted to set metrics that would be challenging and inspiring to us as individuals—but also ones that, if met, would lead to real change on the ground with our graduate students. Setting a target that 100% of faculty would pass the SGA assessment was a lofty goal, but we knew that we didn’t want to leave a single faculty member behind. After our initial improvement cycle, we knew we should keep pushing forward with our faculty training and support efforts and that we should continue to develop improvement strategies for the goals where we fell short.

It was at this point that we knew our improvement efforts were not over. We’d made some great headway, particularly with respect to faculty perception of their training and support, but we knew we had more work to do both to improve graduate students’ experience with Relay GSE data trackers and to improve the fidelity of implementation among our faculty. We continued our improvement efforts over the course of the following school year, creating three new improvement initiatives in order to meet our six quantitative improvement goals.

**Improve: Round Two**

**Improvement to Products**

**CONTINUE TO IMPROVE TRACKER FUNCTIONALITY**

Unfortunately, the small fixes we made to our data trackers the previous year had not resulted in a sufficiently large improvement in graduate students’ experience with those trackers. The initial feature changes we made to the tracker may have improved the experience for advanced users, but it didn’t reduce the learning curve for novice users. At the time we surveyed our graduate students, only 55% of students agreed or strongly agreed that they knew how to use their data trackers. This was a real problem, as all of our graduate students were required to use these tools in order to complete their SGA coursework and graduate from the program. Feedback from faculty and graduate students indicated three major, nagging problems:

1. The data trackers were such large files that they would often crash users’ computers. This technical malfunction required many faculty members to directly assist graduate students with inputting and quality assuring their data, as opposed to their doing so independently.
The data tracker was protected to prevent mistakes, but these protections also prevented users from performing simple operations like copy/paste within the data tracker. These features helped to increase data quality, but they confused users who were not already familiar with data validation and restrictions in the input fields.

Certain types of user error would permanently corrupt the tracker functionality, at which point graduate students were required to start anew with a blank data tracker. This malfunction would prompt incredible frustration for users, causing them to lean on others for help and support, or to limiting their willingness to continue working in order to fully learn how the tracker functions.

Knowing that we needed support beyond our personal skill sets and expertise, we partnered with Relay’s technology team for a solution. The technology team programmed a set of macros within the data tracker that then allowed graduate students to update their data without disruption and which had the added benefit of providing much better quality assurance of the data. The macro functionality provided more checks to prevent errors from being introduced into the tracker while also removing many of the previous obstacles to efficient data entry. Overall, the introduction of the macros succeeded in stabilizing the tracker tool and improving the data-management experience for graduate students. Unfortunately, this was not a perfect solution, and in the lessons learned section of this chapter we discuss the limitations of this improvement effort.

We also knew we needed to do more to train our graduate students to use the tracker tools more effectively. Within our SGA courses, we had online tracker tutorials that taught graduate students how to store and explore their data. Given the updates to our standards mastery tracker, we knew this was also a great time to revamp the online trainings for that tool. Learning from our experiences with faculty members, we wondered whether our graduate students would also benefit from a cultural change effort. While we had worked hard to change faculty members’ attitudes toward the trackers, we realized we hadn’t translated this change into our graduate student-facing materials. In response, we rewrote all of our online tracker tutorials and totally changed the tone of those trainings. Rather than focus merely on having our tutorials be straightforward and clear, we infused humor, jokes, and pop-culture references into the tutorials and tried to lighten up the experience for graduate students. Right when they might be expecting one of the most dull or frustrating aspects of their SGA work (using Excel spreadsheets), we surprised graduate students by making the work fun and amusing. Updating the online tracker tutorials for graduate students had the added benefit of improving our tracker training for faculty the following
year, as we dual-purposed much of our student-facing SGA materials for our SGA faculty training.

Improvement to Training

Created Online SGA Course for Faculty to Bolster In-Person New Faculty Training

After the culmination of our first SGA faculty assessment, 77% of faculty members who took the assessment passed on the first attempt. After additional coaching and supports, nearly 100% of faculty passed on their second attempt. These data let us know that we needed to provide additional supports for faculty beyond our two-day new-faculty training. In response, we created an online course for faculty to take after completing SGA faculty training. The online course consisted of more scaffolded practice for faculty on three topics: SGA measurement principles, tracker use, and the five-step measurement process that we asked our graduate students to follow. These practice opportunities allowed faculty members to have an additional opportunity to step into the shoes of our graduate students, as well as to have additional low-stakes at-bats with the SGA work they would be leading as faculty members come the beginning of the school year. After completing these additional rounds of practice, faculty members were given access to answer keys for the simulated activities, enabling them to check their work and go back and address any misconceptions they may have had. Then, when individual faculty members felt ready and prepared, they could take the SGA faculty assessment at a time of their own choosing within a given time period. In contrast to the previous year, when faculty members all took the assessment at a similar date and time, faculty members could now take any additional time they needed to study and prepare for their teaching duties in the fall. During this assessment window, we also provided SGA office hours for any faculty who wanted additional one-on-one support. The SGA assessment asked faculty members to do things like answer graduate student questions about their SGA pathways and fix broken trackers. It was our intention to have this assessment mimic the real work of teaching and supporting SGA coursework.

Improvement to Fidelity of Implementation

Improved Score-Norming Resources

Our experiences to date had led us to believe that, if we provided faculty members with the appropriate resources and support, they would make the SGA content sing. When we looked at our SGA course-assessment audit data, where we only agreed with assessment scores in an average of 70% of the assessments we audited, we realized we could be doing more to provide
faculties with resources to help them score more accurately from the get-go. While our post-scoring feedback seemed to be helpful, we wanted to provide better resources to faculty before they scored an assessment. In response, we added two new resources to every SGA course we designed: first, we created an online score-norming opportunity for every course assessment; second, we created exemplars for course assessments so that faculty members would have an assessment that could serve as an anchor for their scoring.

**Improve: Results of Round Two**

At the end of our second improvement cycle, we collected a new round of data to see where we stood with regard to our six improvement goals. After collecting these data, we were excited to see that we had met an additional two goals and made improvements in the two remaining ones—making a 21% improvement gain for our tracker-use goal and coming within 1% of our tracker-support goal. While we knew we still had work to do to improve our data trackers and to increase their ease of use, we felt confident that we had made substantial changes to the quality of our training and support of faculty members, as well as to the perception of the SGA content in our graduate students and faculty alike. Our improvement efforts could now be deemed a success by the quantitative measures we set, and our conversations with faculty members and observations of their teaching only further confirmed that far more faculty members really did love SGA than had in years past (see Table 7.2).

**Control**

The control phase of the DMAIC model focuses on ensuring the stabilization, reliability, and sustainability of the improvement activities completed in the prior phase (Basu, 2004). This phase is the end point for the DMAIC process. The control phase helps to ensure that the improvement efforts have taken root in everyday operations and activities and that checks and balances are in place to support the permanence of the improvements. In implementation science, the control phase mirrors the “integration” stage, which describes the point at which a change has been fully taken up and folded into standard practice (Dowrick et al., 2009).

One of the most exciting challenges we faced in the control phase was meeting the demands of an organization undergoing rapid growth. Relay GSE expanded from a single campus in New York to five campuses across the country within just two years. During that time, our work to improve the SGA curriculum broadened to encompass a large number of new faculty who were hired and many new regional contexts that were established. We were not just controlling the improvements for the existing faculty but were also working to ensure that new hires would experience those same
Table 7.2 Round 2 Progress toward Goals

<table>
<thead>
<tr>
<th>Goal</th>
<th>Results Round 1</th>
<th>Results Round 2</th>
<th>Met?</th>
</tr>
</thead>
<tbody>
<tr>
<td>80% of graduate students Agree or Strongly Agree that they know how to use their data trackers</td>
<td>55% of graduate students A/SA (n = 273)</td>
<td>76% of graduate students A/SA (n = 615)</td>
<td>No; 21% improvement</td>
</tr>
<tr>
<td>80% of graduate students Agree or Strongly Agree that they received the support they needed with their data trackers</td>
<td>74% of graduate students A/SA (n = 326)</td>
<td>79% of graduate students A/SA (n = 578)</td>
<td>No; 5% improvement</td>
</tr>
<tr>
<td>90% of faculty members will Agree or Strongly Agree that the SGA team provided them with high-quality TRAINING</td>
<td>100% of faculty members A/SA (n = 11)</td>
<td>94% of faculty members A/SA (n = 16)</td>
<td>Yes</td>
</tr>
<tr>
<td>90% of faculty members will Agree or Strongly Agree that the SGA team provided them with high-quality SUPPORT</td>
<td>91% of faculty members A/SA (n = 11)</td>
<td>94% of faculty members A/SA (n = 16)</td>
<td>Yes</td>
</tr>
<tr>
<td>100% of faculty members will demonstrate proficiency of SGA measurement principles before teaching SGA coursework</td>
<td>80% of faculty demonstrated proficiency on first attempt of the SGA faculty assessment (n = 26)</td>
<td>100% of faculty demonstrated proficiency on first attempt of the SGA faculty assessment (n = 17)</td>
<td>Yes</td>
</tr>
<tr>
<td>SGA team will agree with faculty decisions 80% of the time during SGA course-assessment audits</td>
<td>SGA team agreed with 70% of faculty scoring decisions (n = 50)</td>
<td>SGA team agreed with 85% of faculty scoring decisions (n = 47)</td>
<td>Yes</td>
</tr>
</tbody>
</table>

improvements in their engagement with the SGA curriculum rather than passing through the rite of passage that our inaugural faculty had undergone as we worked out the problems with SGA. Much of our strategic approach to the control phase was guided by the leadership and governance structure of
Relay GSE. The organization staffs each regional campus with its own dean, and this person has ownership for the training, support, and development of all regional faculty. These regional deans would ultimately become the quality-assurance linchpins that would control the improvements we made.

For those new faculty, we established structures through which we could continue to train and support their development in partnership with regional deans. We held open lines of communication with deans and kept them informed of new faculty progress with onboarding requirements and performance on SGA audits. All of us were engaged in the improvement process for the same reason—to ensure that our graduate students were most effectively and successfully prepared to measure and drive PK–12 student learning. Relay GSE’s hiring process ensured that this collaborative spirit and alignment around a shared mission was present for all Relay GSE faculty, and this aspect was of critical importance to making our improvement efforts work. In another organization, or with a different faculty culture, it is entirely possible that we might have been perceived as overseers or evaluators of new faculty performance (which was absolutely not what we were). We made a deliberate effort to ensure that the nature of the relationship between team SGA, deans, and new faculty truly became one of partnership, and we owe a great deal of credit to the organizational forces that created the climate so necessary to establishing those partnerships.

We also continued to survey existing faculty about what was working, what could be improved, and how well the changes were sticking. We observed faculty sessions and guest-taught SGA classes. Though we were our own two-person team in our school’s organization chart, we straddled somewhere in between curriculum-design staff and teaching faculty. We insisted on being faculty first, however, and design staff second. We wanted to work with and for our faculty members and to stay as close as possible to the real teaching and implementation of our coursework. Knowing as we did that the success of SGA lay in the strength of our faculty members, we wanted to develop the closest possible relationships with them so that we could continue to keep an open dialogue about what was working and what was not as our school continued to grow and our contexts became increasingly complex.

At the time of this writing, nearly seven years since the formation of the SGA team at Relay GSE, regional deans have now assumed full, autonomous control over the SGA curriculum, the training of their faculty, and the satisfaction of their graduate students. Though the improvement effort we describe in this chapter represents a successful change to faculty training and support with respect to teaching SGA content, regional deans and faculty members have continued to identify aspects of the SGA content and curriculum they want to simplify and continue to improve. While we in no way solved every SGA challenge Relay GSE faced through these targeted improvement efforts, we believe we established a helpful improvement process that Relay GSE instructional leaders may draw on in the future to continue to improve this ambitious aspect of Relay’s academic program.
Lessons Learned

Though this story talks about the creation and rollout of a rather specific set of curricula in a rather specific teacher preparation program, in essence our story is one of innovation and of the challenges that come with doing something new. While not every teacher preparation program may work with in-service teachers nor include SGA coursework like what exists at Relay GSE, all teacher preparation programs are currently charged with continuing to improve their respective programs and with innovating new curricula in response to changes and innovations taking place in the PK–12 space. In light of this, we conclude by sharing some of our most enduring “lessons learned,” which we hope will be helpful to our teacher educator and teacher preparation program colleagues engaging in similar innovation and continuous improvement efforts.

Investigate Root Causes

One of the first suggestions we received when we realized our graduate students and faculty were struggling with SGA content was to have our part-time content faculty teach the SGA curriculum instead of full-time core faculty. The rationale behind this recommendation was that, if we situated the SGA work within the content methods courses our graduate students were also taking, the content faculty would be able to make better sense of our tools and curriculum for our graduate students—in essence, that we’d simply made a structural mistake in organizing our curriculum and that what we needed was a structural response.

However, we knew from our observations and interviews with faculty that the issues they were facing were not structural challenges—they were knowledge, skill, and mind-set gaps. Our inaugural faculty were working their tails off to learn the new SGA content as they were rolling it out to their graduate students, and they didn’t always feel as knowledgeable and comfortable with our SGA content as they were with other content they were teaching. If we had been tempted to take the quick-fix route to our initial problems and simply switched who was teaching SGA content, our experience now leads us to believe that not only would this not have solved any of our problems, but would in fact have been likely to worsen them. Quick fixes are tempting—particularly in an organization as urgent as the one in which we work—but they don’t always lead to the best solutions. While we did take some time to fully understand our problem, and that did “slow us down” somewhat in the beginning stages of our work, not proceeding down the path of an untested and under-interrogated quick fix saved us countless hours—if not countless months—in backtracking a weak improvement plan.

Furthermore, the DMAIC process caused us to challenge our own biases. We enjoy creating sustainable and replicable processes, and we enjoy
making systems that work at scale. However, no system we created was
going to solve the problems we sought to improve in our SGA coursework.
Our problem was a human problem, and it needed a human solution. With-
out systematically interrogating our problem, we could have just as eas-
ily gone down another misguided track. At the end of this improvement
effort, we believed the time we invested to fully understand our problem
and to understand the stakeholders experiencing our problem was entirely
worthwhile.

**There Is No Way Around a Bad Tool**

From the beginning, our data trackers were not efficient and effective tools.
We started with a faulty tool (with the best of intentions) and spent years
trying to build workarounds for what was ultimately a bad product. In
the first stage of our improvement work, we investigated what it would
take to completely retool our data trackers and found that it would be a
costly and laborious process that we could not in good faith recommend
at the time, knowing our other institutional commitments. We also knew
that we did not have the organizational support for a massive overhaul of
our trackers at that time. In response, we created complicated workarounds
including substantial faculty training, support queues, and a lot of hype.
Looking back on the past several years of our work, we’re not certain that
not fixing our data trackers from the very beginning didn’t cost us much
in human capital as it would have cost to develop new tracker tools—i.e.,
we may have spent as much time and money asking our faculty members
to individually fix trackers as we would have spent on redesigning better
ones that required less upkeep and support. According to the Improvement
Guide, “A technology that is unreliable is worse than none at all” (Langley
et al., 2009, p. 127). If we had to do it all over again, we are fairly certain
we would have pushed harder to provide our faculty members and graduate
students with a better tracker tool earlier on in our history as a graduate
school of education. While we are incredibly proud of the rigor and purpose
of Relay GSE’s SGA coursework and of how our deans and faculty have
taken ownership of this work, we believe we may have put too much unnec-
essary burden on the shoulders of our faculty members.

**Continuous Improvement Is about Hard Work, not Lowering Standards**

There are countless examples of solutions that engineer better results but
that actually fail to improve a process. For instance, police departments can
increase arrests without necessarily keeping the streets safer, educators can
limit the number of failing grades without necessarily increasing learning,
and governments can reduce spending without necessarily saving money.
The purpose of a continuous improvement process is not simply lowering
standards in a way that leads to what may be seen on paper as improved outcomes—it is about ensuring that the processes and outcomes are actually both improved. That’s what the DMAIC process espouses, and that’s the approach we wanted to take.

The challenge is that it’s much easier to look for simple solutions that quickly change outcomes than it is to work on continuously improving a process. In our situation, we could have lowered the rigor of the SGA coursework, made SGA an elective instead of required course for graduate students, designated only a few faculty members to teach SGA courses, or allowed graduate students to use any data tracker they liked. These changes would have immediately reduced the workload for faculty and graduate students alike and would likely have improved perceptions among the self-selected group that chose to participate. However, those short-term wins would have led to longer-term problems and would have greatly lowered our standards for teacher preparation. As we look back on the hard work it took to improve our SGA process, we believe our efforts wholly worthwhile or, as we like to say, that “The juice was worth the squeeze.” We are also grateful to Relay GSE leadership for holding the line, and for their commitment to graduating a critical mass of outcome-driven, data-literate teachers.

Conclusion

Though we were pleased with the improvement in faculty and graduate student experiences related to SGA coursework, perhaps our biggest point of pride was the information we gathered from our alumni about their experience at Relay GSE (see Figure 7.3). In 2015, Relay GSE began surveying our alumni to learn about what they had been doing since graduation and to learn more about their perceptions of the program. One of the questions asked in the alumni survey was “What was Relay’s biggest value-add?” Overwhelmingly, graduate students responded that Relay’s biggest value-add was “measuring and tracking student achievement,” which referred directly to their SGA coursework.

We were particularly proud of these data because, while we certainly want our graduate students to have a good experience while they are with us, our ultimate aim, in everything we did, was to train novice teachers who possess the knowledge, skills, and mind-sets to have a positive impact on the PK–12 students they serve. These data provided a strong signal that our improvement efforts had been successful and that our efforts had been worth it. Furthermore, it has been incredibly rewarding to see the few instances when Relay GSE alums have been hired as new professors at our graduate school. These individuals walk onto the job exceptionally ready, willing, and able to lead future novice teachers to develop their own skills of measuring and tracking student growth and achievement.

Deming wisely pointed out that “shouting ‘improved quality’ will not do it. You got to know what to do” (PDQ Systems, 1984). Although the
How do teacher alumni feel about their Relay GSE experience?

Relay GSE Prepared Me To...

- Create a safe classroom: 62.0%
- Effectively plan lessons and units: 65.2%
- Improve student engagement: 71.8%
- Use formal and informal assessment strategies: 74.2%
- Use feedback to improve my teaching: 74.4%
- Measure and track student growth and achievement: 83.2%
- Overall satisfaction: 72.6%

Figure 7.3 Relay GSE teacher alumni perceptions about their Relay GSE experience

Source: Relay GSE 2015 Alumni Survey Data.
education sector is brimming over with urgency, our experience affirms the need to bring the required discipline and patience in order to know what to do. The continuous improvement process is a happy medium that channels all the hunger for positive change and couples it with a deliberate process that works. Continuous improvement is not merely talking the talk; it’s walking the walk of working in and through the problem to get to the other side. That is the beauty and the brilliance of it, and that is why continuous improvement is a process we will continue to use.

Note
1. Kaycee Salmacia was Relay GSE director of student growth and achievement, and Liam Honigsberg the associate director of student growth and achievement.

References


Data- or evidence-informed decisions that inform continuous improvement of educator preparation programs (EPPs) can clearly not be conducted without the collection of timely, relevant data. However, as others in this book have presented, collecting data is really the first step in grounding improvement efforts. This chapter presents the journey that the Mary Lou Fulton Teachers College (MLFTC) at Arizona State University (ASU) has been making to collect more meaningful data on teacher candidates’ progress, to visualize that data to provide more timely information relevant for decisions close to the candidate learning experiences, and to ultimately provide actionable information to multiple programs within the two divisions of the college that contribute to continuous improvement and innovation.

This chapter is laid out in three parts. The first part describes the structure of the educator preparation programs in the college. The second part describes the iTeachAZ data system that includes a data warehouse, dashboard, and mobile application to support data collection in the field. The ways in which three stakeholder groups use the iTeachAZ dashboard are described: administrators who structure and manage field placements; site coordinators who coach the candidates; and the teacher candidates themselves. These data uses describe the purposes of the iTeachAZ dashboard for monitoring and to provide feedback to teacher candidates from assessments. Finally, the efforts for using the data from the iTeachAZ data system for program improvement provide a story of limited success and significant potential for the future.

Educator Preparation Programs at MLFTC

The ASU MLFTC has 27 teacher preparation programs that are both traditional and alternative preparation models and include multiple opportunities for teacher candidates to experience short (less than a semester) internship opportunities, senior semester student teaching settings, and full- and split-year teacher residencies. Programs are offered at both the undergraduate and the graduate level, as well as an opportunity for a teacher
to obtain certification in a fifth-year context. Residency beginning teacher candidacies at MLFTC can be either paid or unpaid experiences. The college had 770 undergraduate and 164 graduate teacher candidates who obtained certification in the 2016–2017 academic year. Faculty at MLFTC have experienced the consolidation of what previously had been three different teacher preparation programs at three different ASU sites in and around Phoenix. The college also has an online teacher preparation program. The college is currently configured in two divisions. Division I, Teacher Preparation, addresses the teacher preparation courses that provide for certification at both the undergraduate and graduate level. Division II, Educational Leadership and Innovation, focuses on initial certification of principals and continued support of education leaders through an EdD program, supports the advanced preparation of teachers for multiple roles in schools and other learning environments, and prepares the next generation of education leaders and researchers through PhD programs.

The iTeachAZ program, first implemented in 2011, was an early teacher residency program. In the undergraduate program, students typically start as a teacher candidate in their junior year with a one day per week internship in each semester that is largely mentored by a classroom teacher. In their senior year, the teacher candidates are in the classroom with a mentor teacher for a whole year. During the senior-year residency, teacher candidates fully integrate coursework, participate in an apprenticeship, and learn within a cohort model while co-teaching in a pre-K–12 classroom at a partner school. Pedagogy courses are taught in district locations to permit attention to the implementation of research-based practices and facilitate the connection of what the candidates are learning to their practices in the classroom. The program includes site coordinators who work with multiple teacher candidates at a school. The extended time with the school allows the teacher candidate and his or her mentor to engage in multiple types of teaching practices together, including co-teaching, mentor- or student-led teaching to whole classes, small group learning sessions, and the opportunities for one-on-one instructional experiences. This chapter describes the iTeachAZ data dashboard in its use in the undergraduate, senior-year residency program, including views for both the junior-year and senior-year experiences.

**MLFTC Data Systems—iTeachAZ**

MLFTC coordination of its programs that have field experiences is mediated through Tk20, a field placement and assessment platform now available through a company that is now called Watermark. Field placement staff use Tk20 to assign teacher candidates to schools and districts and connect candidates to field staff including site coordinators, university supervisors, and mentor teachers from the schools in which they are placed. Students can submit forms online such as applications for the program and
other documentation needed to proceed through their program. Information about course requirements and assignments are included in binders that each student receives for each course in which they are enrolled during the field experiences. Assignments include the performance assessments, which are observations scored with the TAP™ rubric developed by the National Institute for Excellence in Teaching (NIET) (NIET, n.d.), and a professionalism rubric roughly based on the Danielson framework (Danielson, 2007). Some field experience courses also include what MLFTC faculty call a “signature assignment,” which is loosely aligned with the InTASC standards (CCSSO, 2013) that the Arizona Department of Education has adopted. Binders also include demographic information and a link to the InTASC standards. Clinical faculty, mentor teachers, and teacher candidates can input their data on observations and submit their assignments directly into the Tk20 system. In addition, the Tk20 system facilitates collaborative evaluations online among clinical faculty, field mentors, and teacher candidates through the identification of areas of reinforcement (what the candidates does well and should continue) and refinements (areas in which the candidate needs improvement). The longitudinal nature of the Tk20 system supports documentation of the field experience work and contexts by each teacher candidate.

Much of the information about a particular teacher candidate’s progression through a program can be captured into the Tk20 data warehouse. However, what Tk20 does not do easily is to aggregate and display the data by stakeholder level (administrator, site coordinator, university supervisor, mentor teacher, or teacher candidate level). For instance, through Tk20, the administrator at the Office of Clinical Experience (OCE) cannot see the information about the progress each site coordinator is making across all of the teacher candidates and their mentors in a particular program without running a report that requires additional manipulation in order to visualize. He or she cannot see the progress on key observations each site coordinator is making with a cohort of teacher candidates. Nor can he or she see at a glance the extent to which any particular cohort has teacher candidates who are at risk of not completing assessments at the level they need in order to progress to degree completion or certification.

The iTeachAZ data warehouse, dashboard, and mobile application was designed to provide each stakeholder with access to the data he or she needs in order to support the success of the teacher candidates. Built on a PostgreSQL database, the iTeachAZ warehouse receives data extractions from the Tk20 data sets in order to inform the visualizations of the iTeachAZ dashboard. The dashboard is built on a Corda CenterView 4.0 platform, a performance dashboard software with visual business intelligence built-in that collects data from various data sources in the clients’ environment and presents it in Web-based browsers and mobile devices. Using CenterView, data can be visualized in the form of interactive tables and graphs. This data can easily be exported to formats such as PDF or Excel.
The iTeachAZ dashboard provides visualizations of the teacher preparation program data at multiple levels. The administrator at the Office of Clinical Experience needs to be able to visualize the work of the site coordinators, university supervisors, mentor teachers, and teacher candidates in order to ascertain the extent to which each of the support functions are operating as they should. He needs to be able to identify the site coordinators who are not conducting assessments or program reviews in a timely fashion in order to assure that the teacher candidates are receiving the feedback they need in order to know how well they are doing and what they should focus on in order to improve. The site coordinator needs to be able to track the performances of teacher candidates in order to be sure that they are performing at the level that will result in a recommendation for certification, to be sure that they do not have students who are at risk and in need of some intervention, and to identify the appropriate feedback that teacher candidates need as they gain knowledge and skills in the field experiences. Teacher candidates need to see their own data across assignments and courses to be sure that they are staying on track, experiencing the appropriate assessments, and keeping a focus on what they need to do in order to improve.

In addition, the iTeachAZ dashboard provides a communication channel for teacher candidates and their supervisors to track concerns during the field experience. The iTeachAZ dashboard communicates directly with the field experience staff, site coordinators, and teacher candidates if performance assessments have not been conducted, providing the nudge to ensure that all stakeholders stay on track. Email addresses for site coordinators, mentor teachers, and teacher candidates are included on the home pages and repeated in each of the areas of the dashboard that include specifics of the type of data being collected. The dashboard also includes links to the rubrics that are used in the assessments and walk-throughs so that easy reference is facilitated.

Unique to the iTeachAZ dashboard is the iTeachAZ mobile application. This application provides for two services to expand data collection into the Tk20 and iTeachAZ data warehouses. First, the mobile application allows stakeholders in the field experience to collect data from assessments such as a TAP™ rubric scored observation on any device—laptop, tablet, mobile device, or desktop computer in the field—and upload the data when the individual has access to the Internet, thus reducing the burden of clinical staff and teacher candidates who may be working in remote schools or districts where Internet access is not guaranteed. In addition, the iTeachAZ mobile application provides the context through which concerns and issues about a teacher candidate can be collected directly into the iTeachAZ data warehouse. There are nine forms that track a student who has an issue, and information about that issue is made available through the dashboard. The Office of Student Services uses the iTeachAZ mobile application to track who has received a notice of concern and the subsequent resolution of the issue.
Data Visualization

The iTeachAZ dashboard provides visualizations for performance assessments, walk-throughs, and progress reports. The following sets of illustrations provide details about what each level of stakeholders sees at the administrator, clinical faculty, site coordinator, mentor teacher, and individual teacher candidate levels. Visualizations for the ASU Student Support data collected through the iTeachAZ Mobile App are also provided at the administrator level, though site coordinators and teacher candidates can also visualize the data displays.

Administrator Views of iTeachAZ Dashboard

Administrators in the iTeachAZ dashboard are staff from the OCE who oversee the placement of students in their field teaching experiences and monitor the trajectory of their performances throughout the junior-year field experiences and senior-year residencies. Administrator level access to the iTeachAZ dashboard is also provided to staff from the Office of Student Services (OSS), who provide advising services to students throughout their enrollment in the teacher preparation program. OCE and OSS administrators have access to data visualizations for all ASU instructors of clinical experience courses, site coordinators, mentor teachers, and teacher candidates. They can view data longitudinally by semester from the original set collected in Fall 2012. Administrators have access to all of the views of the site coordinators and the teacher candidates.

The first administrator data display shows the current status of work from all of the site coordinators for the 2018 session. The names of all of the individuals whose data views are included in this paper have been masked to preserve anonymity. All of the filters (placement status, semester, campus, program, etc.) are at the most general level. The administrator can quickly determine in the Current Evaluation # column which performance assessment the site coordinator has conducted and what percent of those performance assessments are complete, as the color-coding will quickly indicate when the completion rate is below 50%. When a site coordinator has not conducted the requisite performance assessment at a set time, the administrator (usually the associate director of the teacher preparation program) receives an email that lists which site coordinator is in arrears for specific teacher candidates. The administrator can also see the status of teacher candidates’ walk-throughs and progress reports, as seen in Figure 8.1.

Of special interest is the column labeled #TCs At Risk, which shows the number of teacher candidates who are at risk for either their scores on the performance assessment, for a student support issue or both. The administrator can drill down to a specific site coordinator and see which students are at risk for what specific reason (TAP™ or Professionalism rubric or student services concern) as well as information on the areas of the rubrics on
which the teacher candidate is supposed to be working (reinforcement and refinement), as shown in Figure 8.2.

Finally, the administrator can drill down to the individual teacher candidate level to gather additional information about the candidate, as seen in Figure 8.3. It is significant to note that this page does not provide information about the teacher candidate’s risk in the student support category. In order to determine what that risk entailed, the administrator would have to move to another section of the dashboard and look up the specific teacher candidate.

At the Student Support tab, the administrator can identify the specific site coordinator and teacher candidate to find out the nature of the teacher candidate support area of concern, the nature of next steps and deadlines, and status of the concern, as seen in Figure 8.4.
From the original administrator’s overview page, he or she can also see the number and nature of alerts that the iTeachAZ system has sent to him/her. These issues include both problematic performance assessments where the teacher candidate has scored below expectations (usually at a level of 3 in a 1–5 scale) on one of the eight TAP™ indicators being used or on the professionalism rubric, has had three absences, or has a student services issue that has been identified and filed. This view provides the ability to immediately drill down to the individual teacher candidate level to find out specific concerns and can be set up to see alerts within various timelines to keep the administrator current.

These views show how the administrator (Division Director, OCE or OSS administrators) can monitor the performances of both the site coordinators and the teacher candidates. The use of additional filters can identify how different campuses are performing, and the search function allows the administrator to identify specific teacher candidates by name or identification number.

Figure 8.2 Drilling down to the site coordinator level to see candidates at risk
Source: Image retrieved from the iTeachAZ dashboard, with identities obscured for privacy purposes.
Figure 8.3 Administrator view of individual teacher candidate
Source: Image retrieved from the iTeachAZ dashboard, with identities obscured for privacy purposes.

Figure 8.4 Administrator view of the information reported as a teacher candidate concern
Source: Image retrieved from the iTeachAZ dashboard, with identities obscured for privacy purposes.
Site Coordinators’ View of iTeachAZ Dashboard

Site coordinators have access to the data of all of the teacher candidates and their mentor teachers at a particular school site and can visualize all of the displays indicated above to monitor the progress of their teacher candidates. Of special interest is the ways in which the performance, walk-through, and progress report data can be displayed to support the planning by the site coordinator for additional work with teacher candidates, through whole cohort or individual candidate sessions. For instance, the view in Figure 8.6 provides the site coordinator with information about the areas of the TAP™ and Professionalism rubrics in which the students are doing well and those in which the candidates as individuals and as a cohort need additional assistance. In this example, the site coordinator might structure an experience for the students to work collaboratively on the TAP™ rubric indicator PIC (Presenting Instructional Content) which has the lowest average score for all of the students across the cohort. The site coordinator might schedule a session where pedagogical methods would be highlighted to identify issues the candidates had and provide alternative ways of addressing a particular common instructional strategy. Alternatively, the site coordinator might identify the TAP™ rubric indicator TCK (Teacher Content Knowledge) for which four of the teacher candidates have scores of one, identifying them as being at risk in this area. The site coordinator can identify through notes of the observation the nature of the content the candidate was addressing in his lesson and identify specific resources that the candidate might access to strengthen this content. Alternatively, if the candidates are all teaching in elementary mathematics with common content topics, the site coordinator...
might plan a group session to review those mathematics ideas and processes in order to improve the ways in which the candidates are addressing the topics.

The site coordinator might identify an indicator in the Professionalism rubric in which some candidates were struggling, such as C1 (Keeping General Records), and work with those individuals on common record keeping strategies. Or the site coordinator might look at one of the areas in which the whole cohort was struggling, such as D1 (Communicates Instructional Program to Parents), and work with the entire group on plans for a parent report card night that might be on the horizon at the school site.

Additional views of the Performance Assessment data are structured to provide bar graphs of aggregate performance by a site coordinator’s cohort of teacher candidates, bar graphs of aggregates of specific indicators that have been the focus of specific Performance Assessments to ascertain that all eight indicators are being addressed, bar graphs that compare areas of reinforcement and refinement for the cohorts over different Performance Assessments to see indicator areas that improve or continue to need work, and a display of the potential refinement resources that the candidates were indicated. A final tab in Figure 8.6 (Refinement Resources) is associated with the Professional Learning Library, which includes multiple pedagogical and learning resources that ASU maintains for teachers, both pre- and in-service (https://pll.asu.edu/p/). Similar data visualizations are available for the site coordinators for the shorter observations or Walk Throughs and for Progress Reports conducted by the mentor teachers.

Figure 8.6 Site coordinator view of cohort of teacher candidates

Source: Image retrieved from the iTeachAZ dashboard, with identities obscured for privacy purposes.
The teacher candidate view of the dashboard provides a summary of his or her assessments that includes links to the data for each of the Performance Assessment observation and the Walk Through conducted by the site coordinator and the Progress Report performed by the mentor teacher at the school (see Figure 8.7). The overview provides immediate reference to the areas in which the teacher candidate needs to focus and the indicators of the TAP™ and Professionalism rubrics which have been identified as needing specific attention. The teacher candidate can drill down into each of the assessments to identify specific information that has been provided by the site coordinator or the mentor teachers, as seen in Figure 8.8. For instance, in Performance Assessment 1, the most recent for the candidate, the site coordinator identified two of the TAP™ rubric indicators and one of the Professionalism indicators for reinforcement and refinement.

The candidate’s performance is included so that he or she can see how he or she did on all of the indicators as a reference for future work. Indication of the performance relative to others in the cohort provides a measure of the extent to which he or she is performing at or above the rate of peers,
suggesting how he or she might prioritize a focus on different aspects of teaching and structuring the work. While the teacher candidate had a number of indicators that were at a level of concern, the intent of the feedback is to focus the candidate on one particular indicator at a time. This targeted feedback allows the candidate to reflect more deeply on an area of her teaching without being overwhelmed by the scores on the performance assessment (see Figure 8.9).

The iTeachAZ system also provides the candidate with specific information from the site coordinator collected during the observation that relate to specific behaviors or recommendations. In the case in Figure 8.10, the site coordinator emphasized that the candidate had done relatively better on the TAP™ indicator ACT (Activities and Materials) by providing evidence of
using both a whiteboard and a physical artifact to model the mathematics topic of fractions. However, the site coordinator provided recommendations to the candidate to move past the aspects of the specific activities and materials in order to connect them to the mathematics of the lesson. For the Professionalism rubric, the site coordinator indicated that the candidate had built professional relationships with the mentor teacher and others, though the coordinator did not identify this as being indicator A1 (Relationships with Others in Schools and the Profession). The site coordinator did indicate some areas on which the teacher candidate did need to work for C1 (General Record Keeping), citing the lack of meeting set deadlines in lesson plans for feedback 48 hours prior to the lesson. While these statements of
evidence and next steps are relatively terse and concise, they provide specific information in a medium that can be collaboratively examined to facilitate discourse between the site coordinator and the teacher candidate.

Teacher candidates are required to indicate in one report that they have examined the information provided to them in the iTeachAZ dashboard. However, it is not clear how many candidates actually substantively examine the data and use it to extract useful information about what they are doing well and what they need to improve. As MLFTC seeks to update and strengthen the data system, focus group interviews with teacher candidates are planned to ascertain usage and relevance of the information.

The teacher candidate can view the information collected during walk-throughs conducted by the site coordinator, which are abbreviated versions of the performance assessments that focus on the types of co-teaching that the candidate and her mentor teacher used (see Figure 8.11). Walk-throughs focus on only the TAP™ rubric indicators and provide the opportunity for reinforcement and refinement as well. Finally, the teacher candidate can view evidence from the progress reports conducted by the mentor teacher.

The third assessment that the teacher candidate can view is the progress report conducted by the mentor teacher. This assessment focuses on both

---

**Figure 8.11** Teacher candidate view of walk-through information

Source: Image retrieved from the iTeachAZ dashboard, with identities obscured for privacy purposes.
Figure 8.12 Teacher candidate view of mentor teacher progress report
Source: Image retrieved from the iTeachAZ dashboard, with identities obscured for privacy purposes.

<table>
<thead>
<tr>
<th>Assessor Name</th>
<th>Kzihd Sdz</th>
<th>Assessor Role</th>
<th>Mentor Teacher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment Date</td>
<td>02/16/2018 09:25 AM</td>
<td>Placement Type</td>
<td>Teacher Candidate</td>
</tr>
<tr>
<td>Course</td>
<td>DDE/569</td>
<td>Section</td>
<td>0101</td>
</tr>
<tr>
<td>TAP Reinforcement</td>
<td>Activities and Materials</td>
<td>Professionalism</td>
<td>A2, Showing Professionalism: Fulfilling Professional Responsibilities</td>
</tr>
<tr>
<td>TAP Refinement</td>
<td>Instructional Plans</td>
<td>Professionalism</td>
<td>A1, Showing Professionalism: Relationships with Others in School and the Profession</td>
</tr>
<tr>
<td>Assessment Tool</td>
<td>Progress Report 3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date</th>
<th>Issue</th>
<th>Planned?</th>
<th>Announced?</th>
</tr>
</thead>
<tbody>
<tr>
<td>02/12/2018</td>
<td>Absence</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Co-Teaching Approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mentor Teacher Leads</td>
</tr>
<tr>
<td>Mentor Teacher Leads</td>
</tr>
<tr>
<td>Teacher Candidate Leads</td>
</tr>
<tr>
<td>Teacher Candidate Leads</td>
</tr>
<tr>
<td>TC leads lesson and MT works with Ss</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Co-Teaching Comments</th>
</tr>
</thead>
</table>

Reinforcement

<table>
<thead>
<tr>
<th>Area of Reinforcement</th>
<th>Teacher Knowledge of Students (TKS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reinforcement Evidence</td>
<td>Ss write 1 or 2 sent. using diff. methods of instruction (tracing, rewriting, writing on own) for indiv. support of writing and illustration a pict to match. 2 Ss use 5 colors and 1 uses at least 3.</td>
</tr>
<tr>
<td>Reinforcement Next Step</td>
<td>Cont. to differentiate method (highlight, rewrite, write on own) and content (number of sent., length of sent.) to support indiv. Ss needs and ability.</td>
</tr>
</tbody>
</table>

Refinement

<table>
<thead>
<tr>
<th>Area of Refinement</th>
<th>Presenting Instructional Content (PIC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refinement Evidence</td>
<td>TC explains lesson to SC for clarity! Remember to post a ss friendly objective and provide a visual model (think of L to RT, word spacing, and punct).</td>
</tr>
<tr>
<td>Refinement Next Step</td>
<td>Always post (and communicate) the objective of the lesson... provide a model for Ss to see and hear the process.</td>
</tr>
</tbody>
</table>

Follow Up

None

Figure 8.13 Teacher candidate view of mentor teacher feedback
Source: Image retrieved from the iTeachAZ dashboard, with identities obscured for privacy purposes.
the TAP™ and Professionalism rubrics, but the mentor teacher does not provide a score on each of the indicators. Rather, the mentor teacher identifies the areas of refinement and reinforcement and provides evidence and recommendations for specific indicators. In addition, the progress report provides evidence of the co-teaching approaches that were used during the observation period, as seen in Figure 8.12.

As seen in Figure 8.13, the mentor teacher commented on the engagement of the students with activities and materials and identified how instructional plans needed to be reviewed on Monday to ensure that they were appropriate and well-conceived. The mentor teacher provides positive evidence of a specific professionalism indicator and identifies a way to improve professional relationships.

These figures provide evidence of the ways in which the iTeachAZ dashboard visualizations provide both monitoring information of the teacher candidate as well as support the feedback that the candidate receives. Additional displays provide evidence of how the candidate has performed on subsequent assessments, allowing for examination of the pedagogical issues the candidate has encountered and how those have evolved over time. Progress is an important demonstration of the field experiences in these senior-year residencies, as candidates have to receive at least a score of three on all of the TAP™ and Professionalism rubric indicators at some point during their field experiences.

Programmatic Use of iTeachAZ Data for Continuous Improvement at MLFTC

The use of the iTeachAZ dashboard to monitor the work of teacher candidates and those who support them during their field experiences serves an important purpose for iTeachAZ data, especially given the size of the teacher preparation program. In addition, the affordance for feedback that can be collaboratively examined by teacher candidates with their site coordinator and mentor teacher is an important contribution to the field experience. There are gaps in the data that are displayed on the iTeachAZ dashboard. For instance, “signature assignments” that are artifacts that are connected to the clinical courses as assignments are uploaded into Tk20, but the tracking of those assignments and scores or feedback are not provided in the dashboard. In particular, several clinical courses within programs are beginning to connect a signature assignment within and across several clinical experiences to show emerging knowledge and skills that cannot be demonstrated in a single clinical course. Being able to visualize ratings and feedback about those performances would provide valuable information to site coordinators and teacher candidates.

But information from Tk20 and iTeachAZ has limited availability outside of the administrators, site coordinators, teacher mentors, and teacher candidates. While upper level administrators—the division director, associate division director, OCE director, and the OSS director—and some associated staff
have access to the data and visualizations, program coordinators, faculty who
work on the revision of curricula and programmatic offerings, do not have
such access. These program coordinators and the largely tenure-track faculty
who teach the preclinical courses do not get to see the areas of the TAP™ or
Professionalism rubrics and the levels at which the teacher candidates perform
throughout their field experiences. In addition, the program coordinators and
preclinical course faculty question the extent to which the TAP™ rubric indic-
ators are providing a valid and reliable measure of candidates’ performances.
While site coordinators and supervisory staff receive ongoing training in the
TAP™ rubric, there are no mechanisms in place to determine inter-rater re-
liability on an ongoing basis. Preclinical course faculty and program coor-
dinators question the extent to which the TAP™ rubric is an appropriate
instrument to use to structure candidate observations, given concerns about
reliability. In addition, there is pressure for giving candidates the passing score
of 3 on each indicator that they need to be successful in their field experiences
and obtain their degrees and certification from the Arizona Department of
Education. To the extent that the non-passing scores focus feedback to the
teacher candidate, this practice is proper and important. However, when one
score on a rubric has such heavy consequences, there is an element of “teach-
ing to the test” and not focusing on important aspects of observations that
might be more critical and a stronger learning experience for the candidates.

ASU has a process for program assessment in the University Office of
Evaluation and Educational Effectiveness. The program assessment plans
submitted by each of the programs in the university are intended to measure
whether or not an academic program has achieved stated objectives regard-
ing program graduates in terms of their knowledge, skills, and abilities. In ad-
dition, the program assessment plans are intended to identify programmatic
areas that might be revised to improve the attainment of their objectives. The
focus of this process should be strongly on continuous improvement.

The assessment plan process lays out six steps that programs should fol-
low to create an assessment plan. These steps include identification of the
program mission, goals, and outcomes. It follows these with identification
of measures, performance criteria and sampling to flesh out how data will
be identified and collected (https://uoeee.asu.edu/creating-plan).

Yearly reports are due each October, and revised assessment plans are
due each December. Faculty reactions to the program assessment process at
ASU is somewhat problematic. The process is frequently seen as a formal-
ity, as they receive little feedback about the appropriateness of their plans,
the quality of their data, and the reflections in which they engage in order
to submit the report. The data they supply, which comes from the Tk20 re-
ports, are not considered sufficiently connected to the programs, largely due
to the lack of access to the data by half of the faculty in a program. Clinical
faculty have not had the opportunity in recent years to substantively inquire
about the data they provide, as program coordinators develop plans and
submit reports largely in isolation from the rest of the preclinical tenure-
track and clinical faculty and instructors.
Improving the Use of iTeachAZ Data for Continuous Improvement at MLFTC

The iTeachAZ data warehouse and dashboard is a complex system to capture data about the field experiences of Teacher Candidates and the support they receive from their site coordinators and mentor teachers. One process that is currently underway is to expand the access to the iTeachAZ data to preclinical faculty who are responsible for programmatic assessment and engaged in continuous improvement and system reform efforts at both the program and the whole college of education level. This process is made more complex because the preclinical faculty, who may be responsible for courses in understanding the culturally diverse learner—a course typically taken in term 3 or term 4—would not receive information about a student’s performance around the ways in which he or she engaged diverse learners until term 7 or term 8. Other faculty who teach the methods courses during the senior-year residence terms are not the instructors of record for the student teaching courses, and access to information about students not assigned to their courses is problematic from a university perspective. We are working on identifying ways to provide these faculty with dummy access to field experience data, but that work is still in the design phase.

The iTeachAZ data warehouse and dashboard are still a sophisticated mechanism to collect information during the field experiences of students in their placements of which we are aware. Efforts are underway at ASU MLFTC to figure out how to share the iTeachAZ process and coding with others through the Ed-Fi Alliance so that others can use the programming to develop alternative data visualization systems. In addition, the college is participating in the UPD Consulting Teacher Prep Data Model and Collaborative Dashboard project (www.updconsulting.com/teacher-preparation-data-model/) to expand the visualization of information the college collects from recruitment to performance of students in the classrooms of teacher candidates and program graduates.

ASU MLFTC is also engaged in substantive reform of the educator preparation programs in the college, a discussion of which is beyond the scope of this chapter. As those efforts mature, a different data collection and visualization system will likely be needed. We welcome the challenge.

References

The press on teacher preparation programs to become users of data is being felt from nearly every sector of education, ranging from policy and research to practice. Despite the growing urgency, preparation programs may struggle with how to respond, in part because too few successful models are available. Certainly, there are unique qualities and challenges for each program attempting to become data informed, but several themes seem to be detectable among those who have made progress toward transformation.

The teacher-preparation program at Texas Tech University is, on the surface, very similar to other programs in state universities. Historically our “brick and mortar” programs, collectively known as Tech Teach, graduate approximately 450 candidates each year, mostly at the elementary level, and the majority of candidates are white females. Over 60% of these candidates return to teach at locations in other parts of the state, most of which are over 300 miles away from the university. This part of the preparation program, however, has strong, close relationships with local districts (i.e., districts in proximity to the university), and preparation faculty meet regularly with district-level administrators, building administrators, and mentor teachers.

Less traditional looking is the distance portion of our preparation program, which graduates approximately 150 candidates each year. This part of the program, known as Tech Teach Across Texas, is based in 14 large urban and small rural areas of the state. This program also consists almost entirely of female candidates preparing for elementary certification but who represent a much more diverse set of races and ethnicities. By design, all of these candidates remain in the districts where they are prepared as part of a district-based, grow-your-own partnership with the university aimed at cultivating and sustaining a diverse teacher-talent pipeline. Preparation faculty also meet regularly with representatives of each district in a manner similar to the “brick and mortar” programs.

This chapter describes important lessons faculty and district partners of the Texas Tech University program have learned through our own efforts to become users of data. Although these lessons are articulated in hindsight, they may provide a “bootstrap” for those who are ready to respond and confirmation for those currently engaged in the struggle. In this chapter, we
will identify five lessons learned on our way to building a data use culture and some of the experiences that provided these valuable lessons.

Lessons Learned

Lesson One: Partnership Helps us Remain Committed to Improvement

Most discussions of data in the context of teacher preparation bemoan the ready access to data that is virtually meaningless for program improvement. More often than not, the data that are available are probably intended to address state compliance-oriented needs. Although compliance is a valid and binding goal for educator preparation programs, cultivation of a data culture requires leadership and rank-and-file commitment to the idea and practice of continuous improvement.

It is a risky but freeing proposition to commit oneself to continuous improvement. It is the acknowledgment that programs could be better; it is the willingness to break from the status quo of hiding behind norms and history. Acknowledgment that programs could be better is something to share with partner districts, because they knew this reality before we shared it. Hiring our graduates, partner districts are fully aware of strengths and weaknesses in our programs and graduates. We learned that our partners were encouraged and excited by our acknowledgment. It was a means for hope and new expectations.

District Partners Provided a Touchstone

Even with partnership and enthusiasm, it can still be daunting to know where to start the process of continuous improvement through data usage. How can outcomes be measured? How does one know what are the most important variables to assess in the teacher education program?

At the outset, we approached our district partners. Given that our partner districts hire our graduates, we considered them the best source for laying out the skills necessary for success on day one in the classroom. Our partners told us that our graduates needed to demonstrate competency in three important areas: (a) to teach effectively, (b) to engage and inspire the students they taught, and (c) to move the needle in terms of the student achievement. After some initial false starts, these three outcomes became our guiding light for articulating a “trademark outcome” for our program:

Graduates of the Tech Teach program will be masters of engaging pedagogy, models of effective learning, and contribute significantly to student achievement.

This trademark outcome communicates to ourselves and our partners the commitment we have to addressing the district needs and improving the success of our graduates. The trademark outcome became a rallying cry and a
way of articulating for ourselves what sets our candidates apart from those graduating from other programs. More importantly in terms of building a data use culture, it provided direction for evaluating and redesigning curriculum, operationalizing course and program outcome variables, approaching analysis strategies, and communicating all of this back with our district partners.

**Make a Public Commitment**

It has been said that willpower is increased by making public announcements of one’s goals. In addition to the trademark outcome formulated for use in our program, we made two ambitious and concrete promises to our partner districts to “hold our feet to the fire” on data use for continuous improvement: (a) that mentor teachers’ value-added scores would be higher in the year of sponsoring a yearlong student-teacher than in the prior year teaching alone, and (b) that the second-year graduates’ value-added scores would be higher than the district average.

Of course, these statements set expectations in our partner districts. Although creating some pressure for the program, these expectations set a positive tone for the partnership. In early internal program evaluation studies, we have found that we can partially substantiate with a few of our partner districts both promises—on average, mentor teachers’ value-added scores are higher during the practicum year than their previous standalone scores, and the performance of K–12 students taught by Texas Tech graduates is, on average, equal to or greater than the district averages. Again, although this was risky, setting a goal and making it public provided strong motivation for data collection and use for continuous improvement, and it provided a touchstone for further collaboration with our partner districts.

**Lesson Two: The Process of Becoming Data Users Is Iterative**

Many of us in higher education tend to be perfectionists. Living in the world of theory and concepts makes it especially easy to play with ideas of perfection; however, in implementing data use for continuous improvement, “the perfect is the enemy of the good.” We realize in retrospect that it was naïve to imagine that the road to a data use culture is smooth and wide. There were few scripts for deriving consensus among faculty, and fewer still for motivating colleagues for really hard work.

Since the time when program faculty and administrators first tackled the question about our valued outcomes, there have been numerous starts and changes. The initial pass was motivated (by some) as much by resistance attempting to maintain the status quo as it was (by others) to revolutionize our program. Yet the start, the first articulation of outcomes and the data need to evaluate progress, began a process and set expectations for results and improvement. The valued outcomes and promises to the district
Lessons Learned

provided the focus for internal conversation between faculty and administrators and the context for following up initial attempts with refinements and improvements. Expect that initial efforts will require subsequent refinement. Here are a few of our early challenges.

Measures Should Be Chosen for Program Improvement, not Individual Research Agendas

To measure the outcome of teaching effectively we knew that we wanted a performance-based assessment that could be used as a learning tool over the course of our program. By learning tool, we meant a teaching effectiveness scoring rubric that could be taught to teacher candidates as a programmatic framework. After establishing candidates’ understanding of what effective teaching looks like, we aspired to use the rubric on a continuous, cyclical basis to shape effective teaching. We needed a rubric that was quick and easy to score so that we could provide immediate feedback to teacher candidates while the lesson was still relevant in their minds.

Not recognizing the complexity and scope of the undertaking, we initially adopted the Performance Assessment for California Teachers (PACT) (e.g., https://scale.stanford.edu/teaching/pact/about), the early predecessor of the edTPA, as our measure to assess effective teaching. Although an excellent tool, we soon learned it was not the teaching tool that we hoped. It took an average of nine hours to score one PACT, and we were unable to provide timely feedback to candidates about their teaching effectiveness. After struggling for some time to make the PACT rubric work for our formative teacher candidate shaping purpose, we decided that we needed a better formative measure. After research and exploration, we settled on the TAP instructional rubric from the National Institute for Excellence in Teaching (NIET) (www.niet.org). Although this change met our program needs for providing timely feedback to teacher candidates, it created great agitation among some faculty, who were counting on doing research with the PACT instrument. Remember the priority.

If You Build It, They May not Come

Kevin Costner built the ball park and the baseball angels played. It is not so easy in building a data use culture. After agreeing upon the three outcomes and creating assessment measures, it was assumed that the teacher educators would quickly adopt use of the data to improve outcomes for their teacher candidates. This proved to be a faulty assumption. Not only did teacher educators not use the new formative measures, they continued to rely on anecdotes to judge the status of their teacher candidates. We built the baseball field, but no one knew how to play. What was needed was a routine that clearly communicated the new expectations and provided context for professional development.
To signal that building a data use culture involved new expectations and required well-designed professional development, a new routine was established by the dean. We launched a new semester-by-semester event called “Data Day.” The purpose of Data Day initially was for clinical faculty to demonstrate how data gathered in formal teaching evaluations had helped inform decisions they had made about coaching teacher candidates and suggestions for the direction the program should go.

It is an understatement to say that the first Data Day events where rough and ugly. Initially, teacher educators’ use of data varied greatly. For example, some faculty displayed data about candidate competency but drew conclusions or made recommendations that were untethered to their results. A few were unable to identify any pattern in the data that pointed to areas needing improvement. Still others admitted that scores were inflated in a desire to reward teacher candidates who were “better” than their peers or to position them well for interviewing principals who might look askance at lower course grades on a transcript. There were, however, bright spots consisting of faculty who identified the need to reallocate instructional focus based on areas where scores were lower and areas where scores were higher. Even with these inconsistencies, the early Data Day events provided valuable information about the status of data use in the college. The event was not used to chastise underperformers but to take stock and plan the next level in professional development.

In time, Data Day has become more impressive and actually began to illustrate how data are being used to improve our graduates and programs. Now, Data Day has evolved into a two or three day extravaganza focusing on integrating clinical and course outcomes with district assessment data. The product of each Data Day is explicit, college-level decision making and plans for improvement. Expect that establishing a data use culture is an evolutionary process and use problem solving to move toward the target.

**Partner Districts Ask, “You want us to provide what?”**

Texas Tech University does not place teacher candidates “shotgun-style” across a variety of school districts. Instead, we only place teacher candidates in designated partner districts. This means that cohorts of teacher candidates and student teachers are placed in a given partner district and overseen by a key teacher educator and faculty member known as a “site coordinator.” Partner district status is established through a memorandum of understanding (MOU) that delineates roles and responsibilities of the university and the district partners. In particular, the MOU spells out the terms for data sharing between the two institutions for the purpose of maximizing program and graduate effectiveness.

The impact our candidates have on student achievement during their yearlong student teaching, and subsequently in their early years of teaching, remains a central tenant of our program and our partnership with school
Lessons Learned

It is also one of the most elusive data points we pursue as a teacher preparation program in the state of Texas. State-level policies prohibit associating student achievement with individual teachers, and any connection that is made is done at the district level. This has required Herculean efforts to communicate need, refine data requests, and ultimately gather from districts whatever student-growth data might be available.

The cleanest variable (though certainly not without controversy) is the value-added score of the mentor teacher compared with a matched sample and over time, with and without the presence of a teacher candidate. Yet, districts in Texas vary dramatically in what resources they have for documenting student growth, and minimum expectations from the state compound the complexity of demonstrating effectiveness. What we have learned is that gathering achievement data requires a person-to-person connection between program personnel and data teams in districts, characterized by tenacity and collegiality, to ensure that district responses are even attempted at all.

Texas Tech University is collaborating with over 18 school districts across the state of Texas. Each one has agreed to provide data that will allow us to determine how well we are performing relative to our two promises. What we have learned is that although district leadership intentions around data sharing are solid, district capacity for sharing data varies widely. In the early days, and continuing today, it was not uncommon for the district’s data person to ask incredulously, “You want what?” Stick with it; it gets easier even though you may have to plan for different districts using different data systems and database structures for recording the same data.

Dashboards to Nowhere

In the early days, our dashboard systems for teacher educators were not very dependable, and they looked like they were created by middle school students. When we purchased Tableau as a visualization tool, we assumed that what we also needed was “professional-grade” expertise, so we hired a data architect and several programmers.

The hope was that the data architect would get with teacher educators to establish multiple data use scenarios that would eventually drive the functionality of the dashboards. Given the necessary learning curves to becoming a data use culture, this investment may not have been the wisest start. Unless the data architect has a very strong background in teacher education, she/he may not know the questions to ask concerning data use and the best visualization approach to quickly provide answers to standing questions.

After three years, our data systems and dashboards are slowly getting better as culture and expertise with data usage improves. With growing sophistication in our data usage, we are now in a much stronger position to know what a highly functional dashboard should look like.
Lesson Three: To Use Data for Improvement Implies a System, and a System Requires Resources

Teacher educators, especially those working in university-based programs, are accustomed to the rewards and expectations around scholarship and scholarly contribution. One early lesson we learned was that gathering and analyzing data for continuous improvement was an activity that had to be recurring, cyclical, and embedded in the normal work flow. Unlike the research article involving a one-time creation of a data set, the systematic use of data within a program requires the construction of infrastructure and processes to gather data and conduct analyses on an ongoing basis, enabling examination of current and longitudinal trends.

The cultivation of a data use culture is surely enriched by a research perspective, but it is not sufficient in and of itself. For the data use culture to thrive, it requires a steady stream of results that enable meaning making about the program. To provide this steady stream of meaningful data, resources from the department or college must be allocated. Resources include time to conduct the analyses and present resources, personnel to support and execute, regular gatherings of stakeholders to discuss and debate the meaning of results, data portals for gathering data, and perhaps even servers for housing data archives and software for data visualization. No doubt there are systems that vary by scale, but some form of each component will be required.

Back to Basics

Like everything else in this process, selecting the right resource was iterative. Below we describe the bare bones resources needed to get started and some of the missteps we made in discovering this need.

First, some sort of database system is needed to keep track of the outcome variables that will drive the program. Most educator preparation programs have some form of database system, and so it may actually be a question of fully utilizing this existing system to measure the formative outcome indicators that really matter. Initially we created custom, local database and visualization systems, but as we become better informed about what we need, we have started searching for a more cost-effective, broadly marketed database system (e.g., TK-20).

Second, having a data analyst/research methodologist on staff is very helpful when addressing the questions that matter most. A data analyst/research methodologist proved useful in conducting statistical analysis on research questions surrounding the outcome data. Advanced doctoral students with quantitative and qualitative backgrounds have typically been sufficient to address our questions.

Finally, the most important resource requirement is time: the time demands on the dean, the chair, and the teacher educators in the trenches.
Lessons Learned

To ensure that this resource is maximized, it is absolutely essential to start new, highly articulated data use routines and professional development that are backed by reporting structures such as Data Day. We have more to say about this in Lesson Four.

You Don’t Have to Go It Alone

Even with concrete guidance like we are trying to provide in this chapter, it may still seem daunting to launch a data use initiative. Way beyond the demands for financial investment is the question of faculty buy-in and leadership. It takes strong leadership to guide change for improvement, which is challenging with even a few resistant faculty or complaining students. The resulting feeling is that of being all alone in the change process, but you need not be.

The University-School Partnership for Renewal of Educator Preparation (USPREP) National Center housed at Texas Tech University is a coalition of university and partner school districts committed to teacher preparation transformation. The coalition provides intensive, ground-level support for teacher preparation reform, including building a data use culture. Cost of membership is typically supported by local foundations committed to the university’s success. With nine current member institutions spread across the Midwest and South, USPREP is looking to support an ever-growing coalition of university partners.

Lessons We Are Still Learning

Lesson Four: Data Collection Is only Half the Battle—Communication across Constituent Groups Is Essential for Real Data Use

Just as cultures may have subgroups that share a common task but also share a common identity with the larger group, data cultures in teacher preparation units may also be subdivided. Groups concerned with curriculum development may have different tasks than those concerned with clinical practice. Cultivation of a data culture cannot erase these differences, but it can help to bridge differences by focusing on a shared outcome.

Familiar Subgroups in Teacher Preparation

What are the subgroups that make up our collection of teacher educators? In the clinical domain we have site coordinators who are primarily responsible for shaping the clinical competencies of our teacher candidates. Site coordinators shape clinical competencies with the cyclical implementation of the “POP cycle.” POP is an acronym for “pre-conference, observation, and post-conference.” Site coordinators and teacher candidates join together
to prepare for, to deliver, and to immediately evaluate the effectiveness of
teacher candidate instruction. The POP cycle occurs among the site coordi-
nator, teacher candidate, and mentor teacher four times during the yearlong
student teaching residency. During Data Day, site coordinators report on
POP cycle data (e.g., candidate scores on individual rubric indicators, use of
student data, and self-reflection during conferences) and trends across their
teacher candidate cohort and on interventions used to increase areas of low
performance.

On the coursework side, anchor faculty are typically tenure-track faculty
with strong background experience and in content-area disciplines (e.g.,
math, English language arts, science, social studies, special education, bi-
lingual education). These anchor faculty lead a group of instructors using
results derived from a design-based research (DBR) methodology. With this
methodology, courses are broken into four or five modules, each with a
learning outcome goal. Data collected from teacher candidates include a
pre-and post-knowledge/reasoning assessment pertaining specifically to the
content of the module. In addition, teacher candidates provide a video of
themselves exhibiting targeted instructional skills while working with small
groups of K–12 students. This video is scored on the extent to which teacher
candidates teach concepts correctly, age appropriately and with the appli-
cable academic language. In the design-based research process, teacher can-
didates’ results iteratively lead to annual modification of the module content
and pedagogical approaches. During the Data Day, anchor faculty report on
trends in teacher candidates’ mastery of coursework content and teaching
as well as the modifications they have made in course modules to improve
teacher candidate results.

Communication across Subgroups Is Critical

The key and challenge to bridging across groups is communication, which
we are accomplishing through the Data Day events. Still, we had evidence
that subgroups were remaining in their own silos, so in January 2017 we
held two consecutive data days—the first concerned with clinical outcomes,
and the second concerned with curriculum outcomes. This time, however,
leaders from each area (clinical and coursework) participated fully in pre-
senting and hearing results. The discussion that occurred during those days
motivated current efforts to better integrate clinical and coursework beyond
our early innovations.

After listening to each other during the Data Day events, teacher educa-
tors learned that teacher candidate growth was being hampered by lack of
integration across their teacher education silos. This has led to integrative
efforts across content-area disciplines (e.g., special education and English
language arts) and functions (e.g., clinical experiences and coursework ex-
pectations). Early integration has been remarkably fruitful, for example,
faculty from special education and English language arts working together
to concretely demonstrate to teacher candidates how to effectively differentiate instruction during small-group reading instruction. Having a solid clinical shaping structure in our teacher education programs, we expect that, in the future, the integration across teacher prep disciplines and functions will lead to the next revolution in teacher preparation transformation.

Lesson Five: We May never Fully Arrive, but the Process Has Opened Our Eyes

As Dean of the College of Education, wherein resides the teacher preparation program, I see evidence of a strong and growing data use culture. I see it in our budget, in our common language, and in our engagement with school district partners. I even see it slowly making inroads into our graduate-level programs that started, like Lesson One, with the articulation of outcomes and the engagement of partners. Despite this progress, there are significant obstacles ranging from demand on resources to convincing those who do not yet embrace the data use culture or share the vision.

Our links outside of the college, however, help us resist the equally strong pressure to return to the old status quo. Our district partners expect evidence of candidates’ effectiveness, and the vision shared with partner institutions in the US PREP coalition help us keep focus. Our bonds within the program also create expectations and a determination to live by the expectations of a data use culture. In the end, embracing a data use culture is about the freeing exhilaration of honestly acknowledging program strengths and weaknesses and working together to improve outcomes for our programs and graduates. It is what we teach our teacher candidates to do in their classrooms; we are simply attempting to “walk our talk.”
Part IV

Conclusions
10 Challenges, Opportunities, and Next Steps

Ellen B. Mandinach and Edith S. Gummer

Every change enterprise encounters both challenges and opportunities, what we refer to here as the CHOPs. Challenges always balance the opportunities. In many instances, one overarching opportunity outweighs the many challenges, even though the number of challenges may be daunting. The overarching opportunity provides the motivation for change. It stimulates institutional growth and organizational learning, whereas the challenges provide the reality check. In the case of data use in educator preparation programs, the opportunity is for programs to use data to inform their practice and to stimulate continuous improvement in the delivery of curricula and practical experiences, aligned to the emerging needs of the United States educational system.

We ground our discussion of the CHOPS in the notions of systems thinking, noting the interplay among the complex and sometimes competing organizations internal and external to educator preparation programs (EPPs). The chapter synthesizes lessons drawn from the previous chapters that address policy, research, and practice. We lead with the positive; that is, the opportunities that can be gained through data use in educator preparation programs. We draw on the many challenges that have been noted through our own work and experience, the challenges cutting-edge programs have faced, and those raised in the research and policy chapters. This confluence enables us to extract processes that are being shown to be effective and those that continue to plague programs. The chapter attempts to generalize how these processes may be implemented in other institutions, noting their diversity and the complexity of such an enterprise. The chapter concludes with a future orientation, an action plan, a delineation of the actionable steps based on the use of data that educator preparation programs can implement to create data cultures that support evidence-based continuous improvement.

Recap of the Book

The chapters in this book are diverse. We have sought chapters that provide varying perspectives of the field from the research, policy, and practice venues. In Chapter 2, we have tried to provide a landscape of how data are
opportunities that could be put in place. In Chapter 3, Allen and Coble provide a detailed examination of state policies. They raise essential concerns about measurement and the appropriateness of metrics and indices being used. In Chapter 4, Peck and Davis provide a scholarly review of the literature. They present a model that consists of tools, people, and organizations and how those components contribute to programmatic improvement. They also identify key components where challenges might occur and how they can be mitigated. In Chapter 5, Weinstein and Anderson present the guiding principles of the Deans for Impact. In describing the quest for a common indicator system, they present examples drawn from some of the organization’s participating institutions. In Chapter 6, Quintana, Richardson, and Reddick tell the story of how UPD Consulting developed the Teacher Preparation Data Model and how it has evolved, using the industry standards of Ed-Fi, to create dashboards and technology solutions for preparation programs. In Chapter 7, Salmacia and Honigsberg tell the story of how the Relay Graduate School of Education has used one metric, Student Growth and Achievement, to examine programmatic improvement. This chapter raises issues of training both faculty and students to achieve reliable and valid data collection. In Chapter 8, Gummer documents how Arizona State University has adopted and developed technologies to support its improvement process. The chapter raises questions about the data being collected and the ability of faculty and staff to use those data to inform their practice. In Chapter 9, Ridley and Hamman present important lessons learned from their experience at building a data culture at Texas Tech University. For each lesson are valuable recommendations of challenges to avoid and positive steps to take. These lessons are both historical and forward thinking.

Opportunities

Educator preparation programs must see themselves as learning organizations where data use for improving practice is central to functioning (Senge, 1990). Opportunities are the heart of the continuous improvement process. They provide the rationale for data use. We have identified two overarching opportunities that motivate data use in the preparation programs: (a) establishing a culture of programmatic improvement and (b) facilitating better alignment and communication with stakeholder groups in a complex and interacting system.

Programmatic Improvement

Programmatic improvement is separated into two parallel but often competing objectives. By far, the most pervasive opportunity for data use is for educator preparation programs to gain insights into how they can improve their programs through an iterative inquiry process that provides them informative, timely, and actionable data on which decisions can be made; that is,
Challenges, Opportunities, and Next Steps

for continuous programmatic improvement purposes. This continuous improvement process helps programs to identify strengths and weaknesses about their programs, their teacher candidates, their graduates, and cooperating components of their programs (e.g., community college feeders and disciplinary departments). It helps programs to understand their impact on teacher candidates, graduates, and schools in which they are being placed.

The second and interrelated purpose is improvement for the objective of meeting accountability measures. There is no question that accountability looms large in a college of education. Therefore, data use enables programs to gain a better sense of the extent to which various accountability requirements, such as state standards, are being met. Data provide insights into the targets for improvement through programmatic evaluations, passing rates of graduates on tests such as the edTPA, Praxis, or PPAT, and credentialing.

Alignment with Stakeholders

Data use provides programs with opportunities to establish, nurture, and maintain improved alignment with various stakeholder groups. First and foremost, the feedback loop between programs and their cooperating districts provides essential information back to the program about the needs of the districts, their students, and the communities they serve. Data collection provides the opportunity to gain better insights into their needs and to understand how well the program is serving those needs. This feedback loop provides a better alignment between curricula and districts. It can facilitate open channels of communication to stimulate changes as the need surfaces. It also helps to lay the foundation for expectations for success.

Data use also provides opportunities for open communication and improvement within and across institutions of higher education. Feedback loops must exist with cooperating educational components. Teacher candidates take content courses in associated university departments. Cooperation must be sought with those department where candidates gain their knowledge of science, physical education, or language arts. Similarly, many educator preparation programs receive candidates from the community college system. This provides an opportunity for better articulation with feeder programs.

Challenges

There is a need for an important and introductory comment to this system—please do not let the daunting list of challenges be a deterrent to the improvement process! No question there are many challenges. Some can be remedied with due diligence; others will take time. We categorize the challenges into five topics: (a) the establishment of a data culture and organizational issues; (b) data, metrics, and indices; (c) technology to support data use; (d) systemic stressors; and (e) the change process.
Data Culture and Organizational Factors

Building a data culture in an educator preparation program is not unlike building a data culture in a school or district. There are some foundational components that are needed, such as using an inquiry cycle, having a vision for why data are to be used, leadership, providing the necessary resources, and having a data system to support the data needs of the institution (Hamilton et al., 2009). Many of the organizational challenges faced by educator preparation programs are created because a data culture has not been deeply embedded into the institution.

It is well known in K–12 that leadership is an essential component in data use (Datnow & Park, 2014; Earl & Katz, 2006). Deans also play a critical role in the establishment of a data culture, hence in part the creation of Deans for Impact. Deans must convey a message of the importance of data use through an explicit vision that is communicated to faculty, staff, and associated stakeholders. Deans also make possible the allocation of resources and incentives to stimulate data use. They make possible the provision for dedicated meeting time around data, such as Data Days, which are mentioned in several chapters. An essential part of the success of Data Days is the creation of norms for collegiality, as Peck and Davis note. Colleagues need to know how to discuss the data. There is a need for the dean or leader of the event to know how to manage dissent. There also is a need for open, trusting, and nonpunitive discussion. Challenges occur when these components do not exist.

Data literacy and the human capacity to use data are significant challenges. Faculty and staff need to know how to use data and for what purposes. This may be accomplished through training or hiring knowledgeable data analysts and data technology staff to manage the data systems. The data users also need to believe that the data use will have outcomes; therefore, negative perceptions, skepticism, and tensions around data use can be real issues. Faculty may well invoke their right to academic freedom. Others may resist change. This is where leadership and vision are essential, through explicit communication about expectations for data use.

Data, Metrics, and Indices

The decisions that educator preparation programs must make are only as good as the data on which the interpretations will be made. Programs have data to inform their improvement process that have certain characteristics. Often those characteristics are lacking. Data must be meaningful, timely, complete, and relevant. Perhaps most important, they must be actionable; that is, some must yield information that can lead to modifications in program practices. Identifying the right data, metrics, and indices has been a long-standing discussion in the field (CCSSO, 2016; Crowe, Allen, & Coble, 2013; Deans for Impact, 2016; Teacher Preparation Analytics, 2016); and
then, having agreed upon and common metrics may be even more difficult, given program diversity.

The metrics challenges pertain to fundamental principles of measurement. What are considered reliable and valid measures? For many indices, has inter-rater reliability been established for observational data? Are there measures that can differentiate different skill levels of teacher candidates, graduates, or programs? As Chapter 3 by Allen and Coble notes, there are fundamental questions about how to measure certain skills like teacher impact and demonstrated teaching skills. The impact question leads to the issue of the logic model that lays out the following: programs prepare candidates who then teach students, with both the programs and candidates being evaluated on achievement. Many have raised concerns whether candidates’ students’ achievement can be considered a realistic outcome or impact measure because of the many confounding factors. Overall, many unanswered issues remain with respect to data, metrics, and indices:

- Do we have the right indices?
- Are there measurement issues with the indices?
- Are there data quality concerns, such as gaps and incompleteness?
- Are the measures valid for the purposes of informing programming improvement?

**Technology to Support Data Use**

Technology solutions to programmatic data use have been evolving but still present challenges to institutions of all sizes. Many institutions simply have poor and inadequate data systems that cannot support the kind of data and data collections needed. Institutions complain of siloed data, meaning that there is no interoperability across disparate systems, and no means of bringing together the needed data in one repository. Technology solutions still have not been able to capture certain kinds of data that can be informative to the improvement process. They are poor mechanisms for data collection. Finally, access to the needed data can be a problem. Access issues can be within an institution or with who is granted access to specific data. Access to other institutions’ data systems also can be a problem. For example, educator preparation programs likely will not have access to the feeder community college systems, nor will they have access to district data systems where there are needed outcome measures of their candidates’ or graduates’ students’ performance.

**Systemic Stressors and Components**

As in any complex organization, different components leverage different emphases and pressures. Educator preparation programs are no different. They must function within a system that includes: (a) state and federal
standards, requirements, and policies; (b) schools, districts, and communities; (c) university-wide policies and structures; (d) other university departments and colleges; (e) the community college feeder system; and (f) other relevant stakeholder groups.

Policies and regulations can be problematic. As is noted by Allen and Coble in their chapter, there is a lack of consistency of policies across and within states and program to program. The policies must be defensible and effective, yet often they are not. They must be transparent, but they tend to be less so. They also reflect the current political agenda, which can create a ping-pong match, depending on changes in political leadership. As Allen and Coble note, the policies and their ensuing accountability systems must be designed to be fair and defensible. Then there is a further challenge in that states often lack the resources and staff capacity to conduct the required reviews, thereby causing them to rely on other sources of data and information on which to evaluate programs. Finally, there is the issue of public shaming and punitive actions. From accountability sometimes comes the public shaming of poor performance or even failure. Such is the case when the National Council of Teacher Quality (NCTQ, 2014) releases its program ratings.

The systemic nature of various stakeholder groups and their influences provides a challenge for the educator preparation programs. Within the university itself, the program must seek cooperation from other departments where the teacher candidates take their content courses. The challenge arises in how to cooperatively and constructively leverage feedback to affect change across departments. A similar issue arises when the program receives many of its students from the community college system as feeder institutions. How do the programs work cooperatively with external institutions for improvement? Because the programs require cooperation with participating districts, there are complex interactions and understandings that must be put in place, beyond the establishment of MOUs. One is an understanding of local priorities and how those needs become reflected in the curricula. A second is the kinds of supports that are needed for the teacher candidates. A third that is essential for the collection of data needed for the improvement process concerns data sharing agreements. The districts need to collect and provide data to the EPPs that are informative and actionable, yet FERPA may interfere. The protection of the privacy and confidentiality of student data is critical, yet the issue concerns the need for data that may be protected and how that is reconciled.

A final systemic challenge may occur within the university structure. If participation in the data for improvement process requires extra work for faculty, it may necessitate changes to job descriptions and ultimately to tenure and promotion decisions. If this occurs within the college of education, there may be ramifications for other departments and colleges within the university that requires consideration by the faculty senate and changes university wide.
The Change Process

The change process is challenged by the fact that it is highly systemic, comprised of interacting but often competing components. If properly aligned, these components should function in concert with one another. More likely than not, they create unavoidable friction that requires political savvy, cooperation, trust, fear from punitive reprisals, and compromise. This refers to cooperation within the program as well as to associated departments, districts, and feeder institutions.

A second challenge within the change process is that programs differ. Because of the diversity across programs, there is no one-size-fits-all solution to programmatic change, the metrics, and the ability to collect and use the data to inform practice. Size makes a difference. Large institutions may be able to afford sophisticated data systems, whereas small departments cannot. Large institutions may suffer greater dissent among faculty, whereas small departments may be more cohesive. Large programs may have greater institutional capacity, whereas small ones may lack capacity. Service area makes a difference. Institutions that are more local and regional may have closer ties to cooperating districts, making MOUs easier to put in place. This kind of cooperation may be a greater challenge for national institutions. Virtual institutions present other challenges because of the diversity of students and their locations. Alternative programs also present different kinds of challenges, particularly when trying to accumulate data about student backgrounds. Residency programs are particularly unique. They may be ahead of the learning curve in instituting programmatic improvement initiatives. Many residency programs have included institutional research and improvement as part of their program structure. Relay Graduate School of Education is a prime example of that (Michael & Susan Dell Foundation, 2016). Because of the diversity, a corollary challenge is fidelity of implementation; that is, to what extent are programs or departments within programs instituting the change process the same way?

Finally, a major challenge is the need for commitment to change and improvement. This goes to the belief system and to the vision for data to inform the improvement process. It also goes to the recognition and acceptance of the fact that change comes slowly. Patience, resilience, and flexibility are needed to make mid-course modifications and be committed to see through the change process. Data-driven decision making and organizational change are founded on the notion of iterative cycles of inquiry. It is an ongoing process of improvement, not a short-term, linear process. It requires a long-term commitment that actually never ends. There is always more improvement and more institutional learning possible—a never-ending cycle.

Next Steps—a Challenge to the Field to Develop the Needed Infrastructure

We have discussed the need to build data cultures in EPPs and the components that comprises such cultures in Chapter 1 and again here among the
CHOPS. Developing, implementing, and sustaining these data cultures is an essential step toward an effective, continuous improvement process. We reiterate here why these components are important, how they need to be implemented, and steps toward sustainability.

Developing and sustaining a data culture requires infrastructure within the EPP and extending to external stakeholder groups. Infrastructure manifests itself in the form of having the right technologies to support data use, building the human capacity to use data, and having the right data to address the questions at hand. Infrastructure requires commitment from EPP leadership, the dean, and department chairs to ensure that all faculty and staff understand the importance of data use in the programmatic improvement process. This can be accomplished through an explicit vision and expectations.

Vision

Having the command commitment from the dean’s office that expressly and explicitly promotes the use of data is an essential step toward building a data culture. Faculty, staff, and other administrators need to understand why data are being used and what impact those data will have. The vision for data use must trickle down to and be mirrored by department chairs. Expectations for data use must be made explicit. Data use must be modeled.

The vision must distinguish between data for accountability and data for continuous improvement. There is no question that EPPs must continue to deal with accountability pressures. Certainly, the data collection process will help EPPs gain a better sense of the extent to which they are meeting accountability indices (i.e., standards, federal and state regulations, and other requirements) and even improve on accountability metrics, such as evaluations, accreditation, and improved passing rates for graduate credentialing. However, the focus on continuous improvement is about understanding the impact of the institution’s program and helping to provide a roadmap toward the improvement of that program. It is about identifying the strengths and weaknesses of teacher candidates, graduates, programs, and components of the programs.

Embedded Culture of Data Use

The administration must provide the foundation for data use. There are some important steps. One step is to provide meeting time for faculty to examine data. As some of the chapters have noted, meeting may take the form of data days or set aside faculty meetings in which data are specifically examined and discussed. A parallel step is to convene groups of faculty as data teams to examine relevant data for departments, specific courses, practica, and the like. A third step is to appoint data administrators or analysts who can assist faculty and staff in the examination of data. A fourth
component is the inclusion of districts, content departments, and feeder institutions into the feedback loop of improvement to find actionable steps that can be taken to modify practice. As chapters in this volume have noted, the interactions with the districts are essential, founded on agreed-upon memoranda of understanding.

A fifth component of an effective data culture is shared respect and open lines of communication, without fear of retribution. Discussions in data teams can often be difficult, with individuals feeling threatened by admissions of less-than-stellar performance. Trust is essential. Faculty and staff must feel safe to openly discuss successes and failures, looking for ways to improve their courses. Faculty, particularly young faculty, must not fear reprisal in terms of evaluations, tenure, or promotion. Without such trust, the data culture can easily dissolve.

Data Systems

Data have become too diverse, too complex, and simply too plentiful for the human mind to handle. Data systems are therefore an important part of the data culture infrastructure to enable the collection, examination, and reporting of the needed data for programmatic improvement. The technological system needs of institutions may differ based on size of program, but the general premise is the same—data need to be collected and stored in some sort of system for future analyses. The fewer the isolated siloes of data, the better. The ASU and UPD chapters focus specifically on technological infrastructure and some possible solutions to very complex data needs.

Building Faculty Capacity

Not every faculty member, administrator, or staff member needs to be a data wonk. They do, however, need to have a basic understanding of why data are important, how they can be used, and the kind of decisions that can be made; that is, they must have at least some low level of data literacy. They need to be good consumers of information. That said, there is a need for some staff to be highly data literate. One solution is to hire data analysts to work with faculty. A concern with having the data analytic capacity located with one individual is what happens if that person leaves. The more the data capacity can be distributed across staff, the better. Further, the more facile with data faculty and administrators are, the more likely they will have a better grasp of the interpretations being made based on the data.

A Systems Orientation

As noted in many of the chapters in this volume, EPPs do not function in isolation. They have tight connections to disciplinary departments within the university where students obtain their content courses. They must
have effective working relationships with the schools and districts where teacher candidates and graduates are placed. They also must have linkages to community colleges that feed students into the programs. This systemic interplay of connections provides opportunities for open lines of communication and alignment with stakeholder groups. The connections should help the EPPs gain a better sense of how to meet the needs of the districts, as well as the needs of the teacher candidates, graduates, and communities more generally. This means better alignment to the local education agencies and the communities they serve. Conversely, feedback about student performance can open communication with feeder institutions and associated disciplinary departments. Such open and frank feedback should lead to better instruction where students receive their content courses. These actions should open channels of communication for programmatic improvement through a process of laying out expectations for performance and identifying actionable steps for program and course modifications.

Final Words

This volume has covered a lot of territory, from policy, to research, to practice. Policy no doubt will continue to drive accountability, standards, and regulations. There is a great deal to learn from cutting-edge institutions. There is also a great deal to understand from the K–12 research on data use and data cultures. There is much we do not know. We do know, however, that creating a sustainable data culture is not easy. For EPPs, this challenge is compounded by the complex systems in which they must function and the subsequent data collection necessary from associated stakeholder groups. It is further compounded by the need to balance data for accountability and data for continuous improvement. There must be a recognition that EPPs must use data to inform their improvement process like any learning organization. Ignoring the data or refusing to recognize the need to improve is no longer an acceptable option. Institutions must learn from both their strengths and weaknesses and determine appropriate courses of action toward improvement. Such a process requires cooperation throughout the institutions. It requires leadership with a vision for data use. The challenges may be many, but the opportunities far outweigh those challenges. The institutions must want to effect this change process. It will take time. It will take a committed effort across faculty and administration. It will require patience, infrastructure, resources, and the willingness to accept change. We can hope to see successful examples as EPPs continue to implement this process.

References

Challenges, Opportunities, and Next Steps


Index

Page numbers in *italics* indicate figures and in **bold** indicate tables on the corresponding page.

academic freedom 80
academic strength 41–42
accountability: data collected for 95–96; policies 38, 96; system 60
accreditation standards 95
adopters 133–136
alignment with stakeholders 203
Allen, M. 11, 35
American Psychological Association (APA) 48
Anderson, C. 95
APA see American Psychological Association (APA)
APA Task Force 45
application programming interface (API) 134
Arizona State University (ASU) 170; building blocks 123–124; Innovation in teacher preparation 28; MLFTC 187
automatic surveys 4
autonomy 80
candidate and completer diversity 42–43
candidate instructional skill 99
candidate selection and completion 41, 66
Carnegie Foundation for the Advancement of Teaching 22
Carnegie’s networked community 22
Carver, P. 16
CCSSO see Council of Chief State School Officers (CCSSO)
challenges 203; data, metrics, and indices 204–205; data culture and organizational factors 204; systemic stressors and components 205–206; technology solutions to programmatic data 205
champions, identify and empower 98, 106
CHOPs 195
CIS see common indicators system (CIS)
CKT see Content Knowledge for Teaching (CKT)
CLASS 107, 108; observation tool 103, 122
colaboration of TPPs 135
Coble, C. 11, 35
Coburn, C. E. 72
Cochran-Smith, M. 26
Colby, S. 21
Collaboration for Effective Educator Development, Accountability, and Reform (CEEDAR) 19
Collaborative Analysis of Student Learning (CASL) 21
colleges of education 3–5, 10, 27, 123
common data collection 100–104

Ball, D. 44
Bastian, K. 76
Bill & Melinda Gates Foundation 118
biweekly SGA digest 155
“brick and mortar” programs 188
Brown, J. 70
Bryk, A. 7, 22
bullying 5
Bunch, G. C. 76
Butler, L. 74
buy-in from faculty 104

CAEP see Council for the Accreditation of Educator Preparation (CAEP)
CAEP-Pearson project 36
common-data effort 98
common data within educator preparation 96
common indicators system
(CIS) 97; data categories and common instruments 100; pilot implementation for featured EPPs 102–103
common indicators system network 98–100
common measures 95–98; common indicators system network 98–100; conditions for collecting common data 100–104; identify and empower champions 106; local priorities 108; Relay Graduate School of Education 104–105; Southern Methodist University 108–110; stakeholders’ values 104; start small 111–112; Temple University 112–113; University of Nevada–Reno 107; University of North Carolina at Charlotte 107–108; University of Southern California 113–114; University of Texas Rio Grande Valley 110–111; University of Virginia 105–106 completer rating of preparation program 45–46
“compliance” mandates 86
compliance motivations 80
compliance-oriented needs 189
consistency of SGA instruction 150–151
contemporary accountability policies 85–86
content analysis 99
content-based courses 5
content knowledge 6
Content Knowledge for Teaching (CKT) 44
continuous improvement 143–146; biweekly SGA digest 155;
consistency of SGA instruction 150–151; created faculty SGA assessment 155–156; data trackers 149; defined 16–17; DMAIC framework 146–149; in EPPs 17–18, 21–22; faculty, stressed and anxious 151; faculty, under-supported 151; faculty knowledge and comfort 149; faculty training for SGA 153–154; fidelity of implementation 160; improvement to training 160; improve tracker functionality 158–160; incremental improvement versus reform and innovation 24–27; investigate root causes 164–165; models 22–24; policies and pressures for EPPs 18–21; SGA course auditing 156–157; small fixes to data trackers 152; tracker-support queues and SGA 154–155
contribution to state teacher needs 48, 67
cooperating teachers 78
Corda CenterView 4.0 172
co-scoring teaching performance assessment portfolios 84
costs for implementation 134
Council for State School Officers (CCSSO) 18–19
Council for the Accreditation of Educator Preparation (CAEP) 19, 36, 69; accreditation of EPP 50; standard 5, continuous improvement 19–20; standards 119
Council of Chief State School Officers (CCSSO) 37, 68
course-assignment audits 150
created faculty SGA assessment 155–156
cross-divisional engagement 73
cross-institutional analysis 97
cross-institutional indicators 114
cross-institutional inquiry 104
cross-institutional learning 99
Crowe, E. 18
curriculum: developers 17;
technology in 5
Curry trailblazing team 105
Cuthrell, K. 88n1
Danielson Framework for Teaching 48
data, metrics, and indices 204–205
data access 134
data analytics 4
data and decision making: in business 4; hospitals 4; in medicine 4; in sports 4
data categories 99
data collection 13, 100; protocols 97
data culture and organizational factors 204
data culture in educator preparation 7–10; academic strength 41–42; accountability 55–58; candidate and completer diversity 42–43; candidate selection and completion 41;
challenges 51–52; completer rating...
of preparation program 45–46;
data and measures 54–55; data on
teacher preparation program 36–38;
demonstrated teaching skill 47–48;
differentiation in performance 55;
diversity of program characteristics
52–53; entry and persistence in
teaching 48–49; general teaching
skill 45; implementing the system
51; inevitability of imperfection 52;
knowledge and skills for teaching
43; K–12 student perceptions
48; limited state resources 54;
mastery of teaching subjects 43–44;
performance as classroom teachers
46; pilot phase 55; placement/
persistence in high-need schools/
subjects 49; political impacts 53;
program performance data 39–40;
program performance indicators
40–41; program performance
measures 50; risk of marginalization
52; state role 35–36; state teacher
needs, contribution to 48; subject-
specific pedagogical knowledge 44;
teacher impact on K–12 students
46–47; teaching promise 42
data dashboards 69, 87
Data Days 79, 192, 195, 204
data-driven decision makers 8
data-driven education 121–122
data gurus 10
data-informed program improvement 18
data literacy 204
data-literacy content 144
Data Quality Campaign 5
“data retreats” 79
“data summits” 79
data systems 70
data tools 75
data trackers 149
data use culture in teacher education
188–189; data collection 195–197;
data for improvement 194–195;
growing data use culture 197;
lessons learned 188–197; partnership
189–190; process of becoming
data users 190–193
data use in teacher education programs
68; accountability policy 68–71;
coda 87–88; faculty engagement
80–82; organizational policy and
practice 71–72; organizational
supports 78–80; people 71; policy
dilemmas and directions 85–86;
problems of practice 82–85; research
72–75; research needs 86–87;
tools 71, 75–78
data use policies 11
data visualization: Educator
Preparation Programs at MLFTC
170–171; iTeachAZ 171–174;
iTeachAZ dashboard 174–178;
iTeachAZ data for continuous
improvement 185–186, 186–187;
site coordinators’ view of iTeachAZ
dashboard 178–180; teacher
candidate views of iTeachAZ
dashboard 181–185
Davis, S. C. 68, 77, 79, 81
Deans for Impact 11
decisional capital 21
decision making 78
deliveryology 22, 23
deming, W. E. 166
demonstrated teaching skill 47–48
design-based implementation research
(DBIR) 22, 23
design thinking 125
Devlin-Scherer, R. 76
diversity of program characteristics
52–53
DMAIC framework 146–149
Duckworth team’s Grit Scale 42
Duguid, P. 70
Duncan, A. 4, 24, 25
Duty, L. 25
Earl, L. M. 9
easy-to-understand format 125
Ed-Fi Alliance 187
Ed-Fi data model with TPDM
extensions 130
Ed-Fi Data Standard 12, 120–122
Ed-Fi for teacher preparation 124
Ed-Fi K–12 core data standard 122
Ed-Fi TPDM 118
EdPrepStat with colleges of teacher
preparation 122–123
edTPA 203
education, process and structure 4
educational data-driven decision
making 7
educational technology 5
educator preparation: data culture
in 7–10; parallel principles in 8
educator preparation programs
(EPPs) 3, 5, 8, 35, 69; continuous
improvement in 17–18; diversity
of institutions 6; institutional
profiles 101; policies and pressures for 18–21; program quality 95; technology into curricula 5
elementary teacher candidates 6
employer feedback 100
English language learners (ELLs) 77
entry and persistence in teaching 48–49
EPPs see educator preparation programs (EPPs)
evidence-based accountability 18
evidence-based continuous improvement 201
evidence-based decision making 85, 87 “evidence-based” practices 82
Evidence-Based System for Teacher Preparation 36
faculty 6; engagement 80–82; knowledge and comfort 149; resistance to change 82; stressed and anxious 151; training for SGA 153–154; under-supported 151
feedback loops 5
fidelity of implementation 160
field supervisors 78
“fire alarm” function 77
focused investigation 87
formal program review 59
formative assessment 5
for-profit programs 6
Franco-Westacott, M. 140
from state-required accountability measures 59
fuzzy stuff 68, 86
Gansle, K. A. 74
Gates Foundation’s Measures of Effective Teaching (MET) study 46
Gomez, L. M. 7
graduate feedback 100
Grady, K. 86
Grunow, A. 7
Gummer, E. S. 3, 10, 13, 16, 170, 201
halo effect 36, 77
Hamman, D. 188
hard-to-staff schools 48
HEA Title II regulations 56
“high data use” programs 79, 81
Higher Education Act (HEA) 36
high-need/high-poverty schools 49
Hironaka, S. 16
homelessness 5
Honigsberg, L. 12, 142
human-centered design 125
implementation science 22, 23
incremental improvement versus reform and innovation 24–27
indicator-by-indicator approach 57
Individual TPPs 135
inevitability of imperfection 52
infrastructure: building faculty capacity 209; data system 209; embedded culture of data use 208–209; systems orientation 209–210; vision 208
inquiry 80
inquiry-oriented program improvement 86
Institute for Education Sciences 4
Institute for Healthcare Improvement (IHI) 16–17
institutional diversity 101
InTASC standards 9
intellectual independence 80
iTeachAZ 13; dashboard 175–185; data warehouse 27, 172; Mobile App 174; program 123
iterative cycle of inquiry 8
Katz, S. 9
K–12 education 6
Kern, T. 25
key effectiveness indicators (KEI) 20, 37, 59, 66–67
Kezar, A. 153
knowledge and skills for teaching 43, 66
Koedel, C. 57
Kolbe, S. 140
Kornfeld, J. 86
K–12 students: perceptions 48; teacher candidates 120
leadership 9
leadership roles and responsibilities 85
lean for education 22
“learner-ready” teacher 19
LeMahieu, P. G. 7
Levine, A. 25
limited state resources 54
local priorities 98, 108
Lubke, P. 88n1
Lys, D. 76, 88n1
Mandinach, E. B. 3, 10, 13, 16, 201
marginalization 52
Marker, P. 86
Mary Lou Fulton Teachers College (MLFTC) 13, 27, 123, 170
Master of Arts in Teaching (MAT) program 113
mastery of teaching subjects 43–44
McDiarmid, G. W. 70
Measures of Effective Teaching (MET) Project 119
memorandum of understandings (MOU) 10, 192
Michael & Susan Dell Foundation 7, 118, 121
Miele, D. B. 72
Missouri Educator Profile 42
MLFTC see Mary Lou Fulton Teachers College (MLFTC)
moment-to-moment actions 4
motivational dynamics 82

Najal, S. P. 17
National Academy of Education (NAEd) 20, 38
national accreditation 36, 59
National Council on Teacher Quality (NCTQ) 4, 206
National Institute for Excellence in Teaching (NIET) 172
National Observational Teaching Examination (NOTE) 44
national professional organizations 11
National Reading Panel 44
networked improvement communities (NIC) 22
Network for Transforming Educator Preparation (NTEP) 19, 37
Nordstrum, L. 16
nursing 72

observation of candidate instructional skill 100
Office of Clinical Experience (OCE) 172
Office of Student Services (OSS) 174
one touch 154
operational data store (ODS) 121, 134
opportunities: alignment with stakeholders 203; programmatic improvement 202–203
organizational supports for data use work 78–80
outcomes-based evaluation 18
Overdeck Family Foundation 118

Pan, Y. 76
Park, S. 16
Parsons, E. 57
Pearson Education 36
Peck, C. A. 11, 68, 70, 77, 79, 81
Penuel, W. R. 23
People-Tools-Organizations (PTO) framework 70–72, 72, 75
performance as classroom teachers 46, 67
Performance Assessment for California Teachers (PACT) 191
performance measures 60
personalized learning 5
pilot implementation 100
placement/persistence in high-need schools/subjects 49
Plan, Do, Study, Act (PDSA) 22, 23
Pointer-Mace, D. 88n1
political impacts 53
POP cycle 195
portfolio artifacts 76
positive deviance 22
positive deviants 24
PPAT 203
Praxis 203
Praxis II assessment 58
Praxis Principles of Learning and Teaching (PLT) 45
preparation programs 11
privacy and confidentiality 10
proactive design 84
process of feedback 8
professionalism rubric 174
program-level data 68
program-level value-added assessment 39
programmatic continuous improvement 3
programmatic diversity 6
programmatic improvement 202–203
program performance 58
program performance indicators 40–41; academic strength 41–42; candidate and completer diversity 42–43; candidate selection and completion 41; completer rating of preparation program 45–46; contribution to state teacher needs 48; demonstrated teaching skill 47–48; entry and persistence in teaching 48–49; knowledge and skills for teaching 43; K–12 student perceptions 48; mastery of teaching subjects 43–44; performance as classroom teachers 46; placement/persistence in high-need schools/subjects 49; subject-specific pedagogical knowledge 44–45; teacher impact on K–12 students 46–47; teaching promise 42
program(s): accountability 68; development 87; documents 85; improvement 68, 78; inspection process 59; leaders 99; performance measures 50; P-12 school partners 85, 87

qualitative analysis of TPA artifacts 76; quality assurance: in education 22, 24; metrics 150; quality improvement 16; Quintana, A. 12, 118

“re-culturing” of education 26; Reddick, C. 12, 118; rehabilitation 72; Reimer, T. 70; Reiser, B. 70; Relay Graduate School of Education 101, 102, 104–105, 142; instructional elements 144; Relay GSE pathway for measuring academic achievement 145; Relay GSE trackers 150; Reusser, J. 74; revolutions for change in education 26; Richardson, B. 12, 118; Ridley, D. S. 13, 188; Rossier School of Education (Rossier) 113; Rossier’s “on-ground” program 113; Ruddell, M. 86

Salmacia, K. A. 12, 142; Scalzo, J. 88n1; Schmidt, W. 44; school climate 5; school of education 5; “school-ready” principals 19; score-norming guidance 157; scoring distribution 39; self-assessment 36; Senge, P. M. 8; SGA Digest 154; Shelby County Schools talent management 123; signature assignment 172; Six Sigma in education 22, 23; Sloan, T. 88n1; small fixes to data trackers 152; social life of information 70; social networks in education 21; social work 72; Solberg, L. 85

Southern Methodist University 101, 102, 108–110; Spillane, J. P. 70, 72; St. John, E. 81; stakeholder engagement 125; stakeholders’ values 98, 104; standard teacher performance assessment 21; Stanford Center for Assessment, Learning, and Equity (SCALE) 44; state data collection 37; state education agencies (SEAs) 118; state educator preparation program approval 37; state implementation of TPPs 135; state licensure 37; strategic organizational support 73; strategic transition planning 85; student diversity 5; student growth and achievement (SGA) 12; course auditing 156–157; curriculum 142, 143; root cause investigation 164; student mobility 5; student services concern 174; subject-specific pedagogical knowledge 44–45; Symonds, M. 74; systemic stressors and components 205–206; Task Force on Educator Preparation and Entry into the Profession 18; TaskstreamTM 21; teacher-candidate development 99, 116n2; Teacher Content Knowledge 178; teacher education programs 68; accountability policy 68–71; accreditation 18; coda 87–88; data use 74–75; faculty engagement 80–82; organizational policy and practice 71–72; organizational supports 78–80; people 71; policy dilemmas and directions 85–86; problems of practice 82–85; research 72–75; research needs 86–87; tools 71, 75–78; teacher educators 35; teacher impact on K–12 students 46–47, 52; teacher preparation analytics (TPA) 11, 20; teacher preparation data model (TPDM) 12, 118–119; adopters
133–136; Arizona State University’s building blocks 123–124; based on best practice 138; better data and data use 119–120; collaborative dashboard 136–137; dashboard design 139; data categories by typical data owner 131; data-driven education 121–122; Ed-Fi data standard 120–121; Ed-Fi for teacher preparation 124; EdPrepStat with colleges of teacher preparation 122–123; inputs to requirements 137; representation 132; Shelby County Schools talent management 123; stakeholder engagement 138–139; use cases 138
Teacher Preparation Data Project 124–125; building the data model 128–129; Chicago Convening, July 2016 125–127; data elements list 128; data owners 131; data systems 131; data systems and owners 129–131; release of 132–133; stakeholder engagement 125; Use Cases 127–128
teacher preparation programs (TPPs) 118; campuses in different states 135; coalition 135; dashboard 139; individual 135; state implementation 135; states, and districts 120
Teacher Work Sample 75–76
Teaching Beliefs and Mindsets 100; Survey 105, 116n3
teaching promise 42
Teachstone’s Classroom Assessment Scoring System (CLASS) 48
Team SGA 144
technology solutions to programmatic data 205
Tech Teach program 188, 189
Temple University 101, 102, 112–113
tenure-track faculty members 103
Texas Education Agency (TEA) 109
Texas Tech University 193
The New Teacher Project (TNTP) 129
TK 20 system 27, 172, 194
total quality management model 74
tracker functionality 158–160
tracker-support queues and SGA 154–155
trailblazing team 99, 104
training, improvement to 160
university-based teacher education 82
University of Nevada–Reno 101, 102, 107
University of North Carolina at Charlotte (UNCC) 102, 107–108
University of Southern California 101, 103, 113–114
University of Texas Rio Grande Valley (UTRGV) 101, 102, 110–111, 138
University of Virginia 101, 103, 105–106
University-School Partnership for Renewal of Educator Preparation (USPREP) 195
U.S. Department of Education 38
value-added assessment model 39
“value-added” programs 74
Vetter, R. 74
virtual programs 6
visualization process 13
Wall, T. J. 74
Web-based dashboards 121
Weinstein, T. 95
Xerox Corporation 70