Data use practices in the continuous assessment and evaluation of educator preparation programs

An Approach to Building Capacity for Data-Driven Continuous Improvement in California State University Educator Preparation Programs

Virginia A. Simon, Sarah Kolbe, and Paul Tuss

Abstract

Educator preparation programs (EPPs) need to ensure that all students have access to effective teachers by using data more strategically to foster improvement. Although many EPPs are committed to improving program outcomes, they often do not use available data to inform their improvement efforts or utilize continuous improvement structures and routines. The California State University (CSU) Educator Quality Center, which serves CSU EPPs, seeks to address these gaps by providing tools and coaching to build improvement capacity. This article describes a year-long effort to investigate EPP data-use practices using an improvement science approach. Examining the roles, routines, and experiences of users gave the EdQ team insight into the conditions that encourage or impede strategic data use and created an opportunity to test promising supports. Three CSU EPPs received structured coaching and facilitation along with resources that helped leaders and teams build trust, consensus, and energy around improvement efforts. This article shares initial findings regarding approaches in these three settings with potential for broader testing and scaling in the future.

Keywords
Data Management, Teacher Training, Curriculum Improvement, Improvement Science

Introduction

The last two decades have ushered in a unique era of accountability for education in the United States (Cochran-Smith et al., 2018). One significant outcome of this accountability era was the creation and expansion of data and assessment systems to measure the effectiveness of teacher preparation programs and the teachers they are preparing. As educator preparation programs (EPPs) have moved to build data systems that comply with new forms of assessment and accreditation, campus resources, habits, and routines around data use have been implemented through an accountability lens. Data collection and use for external accountability are frequently viewed as incompatible with internally motivated data use for improvement by both teacher education practitioners and external audiences (Bullough, Clark & Patterson, 2003). Individual knowledge, beliefs, and assumptions play a major role in how data are interpreted and used in educational settings (Coburn, 2001; Coburn, Toure & Yamashita, 2009; Coburn & Turner, 2011). The too-frequent use of data as a tool for punishment rather than as a resource for learning and improvement has created a defensive stance toward data and eroded trust (Ingram, Louis, & Schroeder, 2004). Furthermore, the available data tend to be aligned to a theory of action that emphasizes aggregate outcomes in performance rather than a more fine-grained view that shows variation, prompting inquiry rather than judgement (Weinstein & Anderson, 2019). This is a tension that is not easily reconciled.

This article describes how the team at the Educator Quality Center (EdQ) is using improvement science, an organizational management approach to problem solving that prioritizes developing, adapting, and implementing reliable processes (Langley et al., 2009;
Bryk, Gomez, Grunow, & LeMahieu, 2015; Dolle et al., 2018), to better understand data use across CSU’s EPPs and to test new ways of engaging with EPPs to nurture a culture of data-informed improvement. As part of the CSU Chancellor’s Office, EdQ has a unique opportunity to target resources and support to each of the 23 CSU credentialing entities. Its mission is to support CSU EPPs by providing system-wide data, evaluation services, and coaching for continuous improvement.

Since 2014, EdQ has worked to expand its operational scope to more comprehensively address the data needs of EPPs. To meet the demand for access to system-wide and campus-level data, it created a data warehouse to integrate siloed EPP data and launched a dashboard reporting system. The first program-perceptions dashboard was released in December 2017 to help CSU EPPs access information on the current needs and perceptions of program completers, new teachers, and their employers regarding the effectiveness of our programs. The dashboard significantly democratized data access, increasing the number of EPP stakeholders able to review the data from less than two dozen to more than 1,000.

While this was an important first step, research demonstrates that simply providing data is not sufficient to ensure its strategic use for improvement (Bryk et al., 2015; Deans for Impact 2019). For example, studies have highlighted collaboration as an additional, critical driver to improvement. A study by Peck and McDonald (2013) of three EPPs found that cross-department, collaborative data conversations that included faculty were critical to faculty-driven improvements in individual programs. Multiple studies have diagnosed that a lack of time to collaborate and a lack of structured routines significantly impedes data use, regardless of data availability (Coburn, 2001; Ingram, Louis, & Schroeder 2004; Means, Padilla, & Gallagher, 2010; Wayman & Stringfield, 2006). Another study found that even when data use for improvement is valued, using intentional collaboration to build understanding and coherence around improvement efforts still requires mindset and organizational shifts (Boudett & City, 2013).

To effectively support campuses, EdQ changed its organizational structure to shift from a data-delivery organization to a delivery and improvement organization. In 2017, they created a data coach position to support effective data use by training end users of the dashboard. Despite the access, exposure, and training provided throughout a one-year period, early monitoring of dashboard usage revealed that while the dashboards were well-received in user testing, only a small number of campuses had more than one or two regular, engaged users. Among those regular users, data analysts and leaders were overrepresented and faculty were underrepresented. EdQ wanted to understand why this was the case. For example, there was no information regarding what collaboration—if any—was occurring around the data. More broadly, EdQ wanted to know how all data sources were being utilized to inform improvement.

**Understanding the Problem**

The team, comprised of a data coach, a data scientist, and the director, began to closely study the feedback and uptake of new users of the EdQ DataView dashboard. They also sought reactions from end users through formal evaluations at their trainings, as well as informal conversations. It quickly became apparent that CSU educator preparation programs are awash in data, not only from EdQ but also their own campus systems, state agencies, and external partners. Furthermore, EdQ learned that the educator preparation community struggles to access, interpret, and regularly use data in strategic ways to improve program outcomes. This discovery is consistent with the findings of a 2011 national study by the Wabash College Center of Inquiry (Blaich & Wise 2011). That study set out to produce comprehensive longitudinal data and measures to provide higher education leaders with data to improve student experiences and outcomes. The design of the study was influenced by three assumptions: a lack of availability of high-quality data was the primary deterrent to strategic data use; providing detailed reports of quality data would spur action; and faculty would engage with the reports in effective ways. At the completion of the project, researchers were surprised to learn that only about 25 percent of the 19 institutions studied engaged in an active response to the data produced and that most circulated data for a short time, with no concrete action or inquiry. This supports prior research on analytics in higher education that found that data use at most institutions was limited to reporting, rather than action, even when a broad range of data was supplied (Bichsel, 2012). The problem was not availability of data, but a gap between access to and utilization of data for improvement of programs and services. The EdQ team chose to focus its improvement efforts on addressing this gap.

**Methods**

In improvement science, all steps are centered around three essential questions:

- What specifically are we trying to accomplish?
- What changes might we introduce and why?
- How will we know that those changes are an improvement? (Bryk et al., 2015).
This 12-month project was organized in phases, reflecting this improvement science approach to clarify and understand the problem and its place in the system (Phase 1), develop a working theory of improvement (Phase 2), and then to test change ideas in rapid iterative cycles (Phase 3). Data are collected during each phase. An Institutional Review Board (IRB) application was not required for this project because it did not involve human subjects. The focus was on institutional program and process quality.

**Phase 1: Understanding the Problem and the System that Produces It**

To explore the causes of a lack of engagement in data use for improvement in CSU EPPs, EdQ conducted a root cause analysis, which clarifies the system of potential causes of a given problem (Bryk et al., 2015) using a fishbone diagram, then highlighting areas where EdQ had some leverage to effect change.

Six broad potential reasons emerged:
1. The organization’s culture of data use is focused on accountability rather than improvement,
2. Lack of awareness of available data,
3. Data are not seen as useful or relevant,
4. Data systems are not user-friendly or intuitive,
5. Data quality or lack of trust in the data, and
6. Misconceptions about how data will be used.

Utilizing existing relationships with EPP leaders and assessment coordinators, the EdQ team conducted a series of data collection activities with an initial set of volunteers to test these potential reasons for a lack of engagement in data use. The team visited one campus to shadow its annual data-review process, and conducted empathy interviews—an interview approach to learn about the experience and feelings of users (Plattner, Meinel, & Lefier, 2014)—to understand what it is like for faculty and staff to use data on the campus. They analyzed the roles of dashboard users to understand how responsibilities for using data are distributed among faculty, leadership, analysts, and other staff members. Team members also met with the deans of education collectively and completed a simplified current state/gap analysis with them about their data-use practices and followed up at a second meeting with a basic root cause analysis to help them and the team understand the barriers to data use for improvement in their settings.

**Phase 1 Findings**

From interviews with leaders, EdQ learned that many faculty and staff have a compliance mindset. One leader reported, “We used to get those big PDF reports from EdQ and we used to have to write a report to our president explaining why the numbers aren’t as good, or the ‘n’ size is low. It was not helpful, just a scary thing.”

They also found variability across campuses in the level of trust faculty have in the data. As one participant stated, “Their first instinct is to discount data—that’s not in-house [data], these aren’t our questions.” An individual from another campus shared a similar observation. “One of my experiences in data meetings is that we spend more time talking about the integrity, validity of the data than talking about how the data can inform our processes.”

In empathy interviews with assessment coordinators, EdQ learned that these staff are often excluded from improvement work and viewed as primarily responsible for compliance reporting. They are frequently asked to provide data without being included in context discussions or follow-up work, which causes them to feel isolated and disconnected from improvement activities. One assessment coordinator shared that, “There’s no formal routine for data meetings. They ask, I do. There are scheduled things that are part of my job—running surveys, producing reports.” Another shared, “Sometimes I go home at night and think, I really hope they’re using this [data analysis] for something… I do wonder what they’re doing with it.”

**Phase 2: Developing a Working Theory of Improvement**

Using the learnings from Phase 1 investigations, the team developed a theory of improvement to begin to focus on what changes might be effective. They determined that potential leverage—or opportunities for high-impact change—existed within three areas. These areas are called primary drivers and are illustrated in Figure 1:

1. Developing leadership capacity to conduct data conversations that promote trust and active engagement of stakeholders;
2. Supporting campus-level routines for regular, collaborative examining of data and testing of solutions; and
3. Establishing system-wide data and visualizations.

In improvement science, the development and testing of changes can take many months. The iterative nature of the data collection and the learning in this phase mean that multiple leverage areas may be examined and multiple changes may be tested simultaneously. For the purposes of this report, the primary testing described relates to the expansion of EdQ’s key role from being an organization that supports EPPs by providing data, to one that partners actively with EPPs to utilize the data to inform...
improvement efforts. Specifically, this report focuses on the new role and activities of the EdQ Data Coach to support leverage areas 1 and 2.

**Phase 2 Findings**

Since Phase 2 constitutes the development of a theory, which provides a structure for moving forward into Phase 3, there are no findings produced in this phase.

**Phase 3: Developing and Testing Change Ideas and Interventions**

Before any interventions could be considered for testing, the team at EdQ needed to define specific high-quality data-use practices and a way to measure their impact. They drew on a resource used in other parts of the United States called the Deans for Impact Data Diagnostic Tool©(2018). The tool is designed to aid teacher-preparation program leaders in assessing the quality of their data-use practices in a collaborative setting. It is a rubric that consists of four domains, each with four to five sub-domains and two to six elements. Each domain can be assessed at a developmental level of “Not Yet Started,” “Emerging,” “Developing,” or “Sustaining.” To simplify the instrument and focus more directly on the goals of EdQ’s project, the rubric was adapted to examine the following domains with no sub-categories:

1. Establishing an inquiry orientation toward the practice of data use,
2. High quality data made available and accessible,
3. Planned and structured collaborative reviews, and
4. Repeated cycles for continuous improvement.

The rubric was used as an internal instrument to establish a baseline measure in cases where the team had engaged more than once with a particular leader or group. It was then used as an outcome measure, showing the progress of campuses that received continuous support from EdQ. It was also considered a potential support tool that might prove as a useful way of engaging campuses in an on-going relationship with EdQ. The aim was to partner with at least two campuses to improve their data use by one level on the rubric within one of these domains by the end of the project.

The team first prototyped multiple protocols for campuses to use the rubric either on their own or facilitated by EdQ as a way to improve data use. They conducted three Plan-Do-Study-Act
(PDSA) cycles to iterate the prototypes with volunteer campus leaders and improvement science experts, gathering feedback through interviews. They learned that the rubric tool was generally positively received as a prototype, but there were mixed opinions about its utility as an exercise for campus groups.

To find out more, EdQ tested a process for facilitating the use of the tool with a campus group. One campus volunteered to host a rubric session with its assessment committee facilitated by EdQ’s data coach. This event was revealing. It showed siloed practices by individual programs, disconnected from a larger strategic plan for the school of education or the university. EdQ’s data coach was invited to return to conduct another session with an expanded group that included faculty program chairs. This group was able to provide more information to the assessment committee about data collection happening within credential programs, surfacing the need to create a more coordinated and cross-departmental system of data collaboration aligned to EPP goals. It also revealed a faculty who tend to view data as an instrument for accountability rather than an instrument for improvement. This developed into a long-term coaching relationship with the dean and associate dean in the months that followed to work toward a coordinated EPP-wide improvement focus.

Being able to deliver support in a virtual setting was an important test for EdQ’s small team in order to effectively and efficiently support 23 campuses across the system. In parallel work, EdQ tested individual leadership coaching and group facilitation supports via the Zoom video-conferencing platform. In order to have a larger sample size for studying the effectiveness of virtual supports, these tests included campus groups and leaders beyond the campuses at the focus of this study.

The first PDSA was conducted to test a virtual group-facilitation approach. Sixty-minute sessions were designed to take groups through guided data digs of the DataView dashboard and introduce concepts and tools for supporting mindset shifts (Domain 1) such as the “ladder of inference” (Senge, 2000; Argyris & Schön, 1974). Data digs showed participants how to look for variation in the data rather than average performance in order to focus their inquiry and choose a problem focus that could lead to an improvement cycle (Domain 4) (Bryk et. al., 2015). The typical group that was gathered by leaders for this session was an assessment committee, task force, or a department chairs group. Leaders were asked to meet with the data coach prior to each session to provide background and context about EPP goals and data-use practices.

Another PDSA was conducted for one-on-one coaching approaches. A coaching protocol was tested to help leaders examine current structures and routines around data conversations, evaluating the type of data viewed, the frequency it is viewed, who views it, and for what purpose (Domain 3). (Sample protocols can be found in this Appendix.)

**Phase 3 Findings**

Data collected during PDSAs included surveys of participants to determine perceived value and relevance of activities to improvement efforts and to assess the value of tools being used. Debriefing sessions were held with leaders to triangulate survey responses and get additional feedback. Detailed notes were kept during each conversation. Leaders were also asked if they would like to receive on-going support from EdQ to find an improvement focus and begin a continuous improvement process with the help of the data coach. Notes were coded to determine leading indicators of progress on the rubric. All coaching and facilitation with campuses was tracked to document the frequency and type of support given. EdQ looked for repeat invitations as a proxy for interest in an ongoing partnership. In all but three cases, they found that the sessions, while perceived as valuable according to survey responses, did not lead to repeat invitations for continuous improvement work. In the three cases where repeat invitations were issued, EdQ was able to form a partnership for deeper, ongoing continuous improvement coaching with the leader and the groups they were convening. These three EPPs, which collectively serve more than 1,200 teacher candidates each year, became the primary sample of the study and are here forward described as Campuses A, B, and C. All three campuses are urban, majority-minority schools.

One theme that emerged across session evaluations was the lack of time. Sessions were typically limited to one hour. When participants were asked what they would change about the session, surveys consistently revealed that people wanted more time to explore their data. EdQ began to ask for 90-minute sessions, but this was not always provided.

**Results and Analysis**

Testing of group and individual support approaches created an opportunity for EdQ to begin to work closely with three EPPs on improving data use for continuous improvement. The results of this early work with EdQs culminated in a total of 22 touch points between November 2018 and May 2019 with either the leader or a group convened by the leader. Two of these ongoing relationships were conducted in virtual settings. One, described earlier, began with an in-person session with the data-use rubric and evolved into a second in-person session and several virtual check-ins with
the leader. The others began with the DataView dashboard session. Following this study, ongoing support has been customized to meet the needs of each leader but generally involves moving a campus team toward focusing on an area of improvement identified in the data and beginning an investigation into root causes and potential areas to focus improvement efforts.

After eight months, EdQ made final designations on the data-use rubric as their outcome measure for the three partner campuses. These final rubric designations were based on structured interviews aligned to the rubric with leaders from the campuses. Two EdQ staff participated in the interviews and calibrated their ratings to arrive at a final determination of whether movement was made on the rubric.

Campuses B and C moved up one level in Domain 1, and campus C also moved up one level in Domain 3 as shown in Figure 2. In each case where improvement occurred, it was from the lowest rubric category (Not yet started) to the second lowest (Emerging). No campuses improved in Domain 4. Campus A did not improve in any domain.

Coaching notes collected throughout the course of the project were coded to identify leading indicators of potential shifts in order to better understand what events led to meaningful improvement on the rubric. Early evidence of change was most often visible in Domain 1, in the form of shifts in mindset as shown in Figure 3. As one associate dean said of their partnership with us, “It’s helped me understand the importance of data. I never really paid attention to
that in the past. I understand the importance of it and how it can be a catalyst for making change in my department.”

Discussion

In this study, intensive coaching and group facilitation by EdQ resulted in movement in two of the three campus partners from the lowest levels of the data-use rubric to the next level. These changes occurred in the categories of mindset shifts and collaborative structures, while no changes were noted in continuous improvement cycles. One conclusion that could be drawn from these results is that these foundational shifts are the easiest to make given the tools tested. Another conclusion is that they are a necessary precondition before actively engaging in an improvement cycle, and therefore making progress on this domain requires more time than the project provided.

This conclusion becomes clearer when considering the following: Campuses B and C had leaders who were new to their positions or new to their campuses. These leaders were eager to make changes in how data were used and how improvement practices were organized. It is no surprise, therefore, that this support paid off in the timespan studied.

Campus A had a leader who was more established and whose practices were already at the Developing level in domain 1 and Emerging in domains 3 and 4. This campus also received the most support (nine events). Where others were just beginning, campus A had already moved, and their task was to build structures to sustain new habits and engage more people. Moving from Developing practices to Sustaining requires more commitment over time and is perhaps the heavier lift for leaders and the system that supports them. New leaders, like those from campuses B and C, will need continued support to sustain the momentum gained. This raises issues for EdQ regarding its capacity to provide this level of coaching to a broader group of leaders in this situation in the future. In addition, with only three out of 11 campuses opting to engage on a deeper level, EdQ will be examining ways to pull more campuses into deeper work with the resources that we have.

While leadership capacity to build trust and guide the culture shift away from compliance and toward continuous improvement is key, systemic supports must be in place in order to sustain the work. Some leaders simply need new tools, processes, and vocabulary to begin. Others require more basic support in how to build consensus and energy around improvement efforts. Still others need structures that reorganize how time and talent are allocated. All of this takes time and focus. For EdQ, this means developing a strategic and differentiated approach to supporting leaders and groups as they use data to improve. Under consideration is the development of cross-campus networks in combination with individual coaching and facilitation support. In other words, EdQ would evolve into a hub for improvement networks focused on system-wide priorities. Additional investigations will examine the role of the assessment coordinator in improvement work and how EdQ can better support them.

Limitations of the Study

There are limitations to this study that are important to acknowledge. Measuring culture shifts and changes in structures and routines during a nine- to 12-month period posed several challenges. Given the time constraint, the scope and sample were small and limited in the depth of detail they were able to capture from each campus. The virtual nature of most of the interactions should also be considered. Too few in-person sessions were held to robustly evaluate effectiveness of in-person versus virtual supports. The data-use rubric as a measure was also limited due to the time needed to demonstrate progress on the practices they were examining.

Overall, the concepts and tools of continuous improvement, used in an improvement science framework, were successful in making some essential foundational shifts in how data were used and interpreted in partner EPPs. Promising practices emerged for supporting these shifts, including developing leadership capacity by providing tools, processes, and vocabulary for building trust and consensus around improvement efforts. Approaches to making the more difficult shifts were harder to measure using the rubric employed. EdQ learned that their partnerships makes a difference. All three EPPs expressed a desire to continue their partnership with EdQ. As one leader stated, “Now I feel like I have a partner in EdQ who doesn’t care about what my score is right now, but about helping me along in the process to figure out how to improve.”

References:


