

SRI International

WestEd 

A System of Measures to Support Improvement in Teacher Preparation

Sola Takahashi
Melissa White
Corey Donahue

2019

© 2019 WestEd. All rights reserved. Permission to reproduce or adapt for non-commercial use, with attribution to WestEd, is hereby granted.

WestEd is a research, development, and service agency whose mission is to promote excellence, achieve equity, and improve learning for children, youth, and adults. For more information about WestEd, visit <http://www.wested.org/>; call 415.565.3000 or, toll-free, (877)4-WestEd; or write: WestEd / 730 Harrison Street / San Francisco, CA 94107-1242.

This publication was made possible by a grant from the S. D. Bechtel, Jr. Foundation via its “Preparing a New Generation of Educators for California” initiative. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the foundation.

The New Generation of Educators Initiative (NGEI) at California State University (CSU), funded by the S. D. Bechtel, Jr. Foundation, seeks to strengthen the current teacher preparation system in California so that new teachers enter the workforce prepared to implement the Common Core State Standards and the Next Generation Science Standards. From January 2015 through June 2019, NGEI is providing grants to CSU campuses and their district partners to improve their teacher preparation programs. The foundation has developed a theory of action to guide reform that focuses on five Key Transformation Elements: partnership with districts, prioritized skills, practice-based clinical preparation, formative feedback on prioritized skills, and data-driven continuous improvement.

WestEd and SRI International are conducting a formative evaluation of NGEI implementation and outcomes at the grantee sites, as well as delivering technical assistance to strategically support data-driven program reform efforts.

Suggested citation: Takahashi, S., White, M., & Donahue, C. (2019). *A system of measures to support improvement in teacher preparation*. San Francisco, CA: WestEd.

Contents

Introduction	1
Background: Continuous Improvement and Process Measures	1
Application to the New Generation of Educators Initiative	4
Partnership With Districts	7
Prioritized Skills	8
Practice-Based Clinical Preparation	10
Formative Feedback on Prioritized Skills	12
Limitations and Challenges	14
Conclusion	15
References	15
List of Figures	
Figure 1. Theory of Action of the New Generation of Educators Initiative	5
Figure 2. System of Measures to Help Guide Teacher Preparation Programs	6

Introduction

As efforts have mounted to reform how teachers are prepared for their profession, so have calls for data that would provide insights into whether teacher preparation programs are producing desired outcomes, and for data that would inform continuous improvement efforts. The New Generation of Educators Initiative (NGEI) at California State University (CSU), funded by the S. D. Bechtel, Jr. Foundation, seeks to strengthen the current teacher preparation system in California so that new teachers enter the workforce prepared to implement the Common Core State Standards and the Next Generation Science Standards. Building on the efforts of CSU teacher preparation programs (TPPs) that have been working toward improved outcomes, this paper offers a perspective on how TPPs can use data that indicate how key parts of their systems are performing.

In the field of teacher preparation, researchers and decision-makers have tended to focus on assessment of the outcomes of teacher preparation. Outcome measures include, for example, the standardized test scores of the eventual students of teachers who have completed a TPP, the employers' perceptions of those teachers' teaching performance, and/or how long they stay in the teaching profession (Deans for Impact, 2016; Feuer, Floden, Chudowsky, & Ahn, 2013; Southern Regional Education Board Teacher Preparation Commission, 2017). TPPs have had less guidance on the kinds of data that would inform continuous improvement efforts. Without the right data, TPPs will not know if their efforts are heading in the right direction as they work toward desired outcomes.

Having data that indicate gaps in outcomes can be a key starting place for improvement efforts, but these types of data may not offer much information to a TPP about which underlying systems or processes within the preparation must change in order to achieve more desirable outcomes.

Background: Continuous Improvement and Process Measures

This paper considers TPPs as systems that can undergo continuous, data-driven improvement through an approach known as improvement science. One of the core beliefs grounding an improvement science approach is summarized in the saying “Every system is perfectly designed to get the results that it gets” (IHI Multimedia Team, 2015). Embedded in this quote is the idea that the processes, structures, and norms that make up systems are what achieve effective outcomes, or are what break down to prevent such outcomes. Rather than placing blame on individuals in the system for undesirable results,

this perspective shifts attention toward the ways in which work unfolds within specific parts of the overall system. When data are used to understand how a system is functioning, multiple measures are needed to shed light on different aspects of the system. Any single measure is insufficient to serve as a representation of the complex and interconnected processes that compose the larger system. A set of multiple measures can provide information about the different system parts in which processes are effective or ineffective in bringing about the intended results.

A set of measures, sometimes called a “family of measures” (Provost & Murray, 2011) or a “system of measures” (Bryk, Gomez, Grunow, & LeMahieu, 2015), includes three types of measures: outcome, process, and balancing.

- *Outcome measures* capture the results of the system and provide information about its overall performance.
- *Process measures* shed light on the functioning of specific parts of the system and indicate whether and to what extent the parts are performing as planned. Process measures are antecedents of outcome measures, and they can serve as early indicators of whether a system is on track to produce the desired outcomes.
- *Balancing measures* become important as people work on improving parts of the system. Balancing measures capture whether efforts to improve one part of the system are having any unintended consequences in other parts of the system.

For TPPs, the outcome measures are connected to the goals of these programs, which typically aim to produce completers (individuals who complete the programs) who are prepared to enact prioritized skills and who will accept jobs in public schools and remain in the teaching profession over time. Thus, key outcome measures might include data about the hiring and retention of those who have completed the program and data about the completers’ teaching effectiveness, such as data from classroom observations, or data from surveys that elicit perceptions of preparedness that are based on completers’ self-reports or on their supervisors’ perceptions of their teaching performance.

The process measures, which are the focus of this paper, are connected to aspects of TPPs that are consequential in bringing about these outcomes. The remainder of this paper focuses on these types of measures for the key components of TPPs. Process measures have several key features that distinguish them from outcome measures. First, process measures illuminate whether specific processes are performing as planned (Bryk et al., 2015; Provost & Murray, 2011). In the context of a focused improvement effort, process measures serve as information about change efforts, as part of a feedback loop, conveying information about whether specific components of the system are, in fact, getting better.

Process measures can also be examined in relation to outcome measures — when process measures show improvement, are there also improvements in outcome measures? They are thus useful, on an ongoing basis, in determining whether adjustments need to be made to the overall theory of improvement. Knowing whether key processes (and potential changes to those processes) are leading to improvements in valued outcomes is a critical use of data.

Second, the data for process measures are typically collected and reported in a more timely manner than the data for outcome measures. Outcomes are often measured on an annual basis, so data on outcome measures may come too late to inform day-to-day improvement efforts. In comparison, data on process measures can be obtained and used in a timely way, usually more frequently than outcome measures (Bryk et al., 2015; Provost & Murray, 2011). More frequent data collection allows for regular monitoring of processes, often as they are unfolding, to enable course corrections midstream. Depending on the process being measured, data may be captured daily, weekly, or monthly. Thus, process measures are leading indicators, rather than lagging indicators.

Third, process measures often address both the quality and the reliability, or consistency, of organizational processes. Those who are involved in efforts to improve systems tend to focus on the quality of system performance. For example, efforts to improve teacher feedback often focus on the effectiveness of the feedback and on whether the feedback leads to improved practice. Although these measures play a critical role in improving processes, an improvement science approach entails examining the reliability, or consistency, of a process, in addition to examining the quality of execution. Measures of process reliability answer questions about whether processes are occurring with consistency over time and across individuals. In the example of feedback to teachers, a high-quality feedback process is not valuable if it does not occur with regularity across the system. Therefore, measures that look at the *frequency* of feedback are needed to ensure that the feedback process is truly high-functioning.

Accordingly, the following sections of this paper focus on *both* the quality and the reliability of given components of the TPP system.

Process Measures Versus Plan-Do-Study-Act Measures

Process measures and Plan-Do-Study-Act (PDSA) measures are easily confused with one another. While there are ways in which these measures connect, they have important conceptual and functional differences in the context of an improvement science effort. Process measures are part of the set of measures that informs an improvement project over the length of time that the project is focused on a particular process. They tend to be captured over the long term, and it is important to establish their connection to the intended outcome of the process and to examine their usefulness for improvement work. PDSA measures, on the other hand, shed light on whether and how a particular change idea works in the context in which it is tested. Depending on the size and scale of the “test,” PDSA measures might be informal and might provide just enough information to suggest whether further testing is warranted. A PDSA measure might be collected just once, to inform one PDSA, or several times over the course of a series of PDSAs that all test one change idea. Although these distinctions exist, the two types of measures can also overlap in use. A process measure might function as a PDSA measure for a particular change idea, if it is captured at the right time, or PDSA measures might turn into process measures as a team settles its focus on a particular part of the system.

Application to the New Generation of Educators Initiative

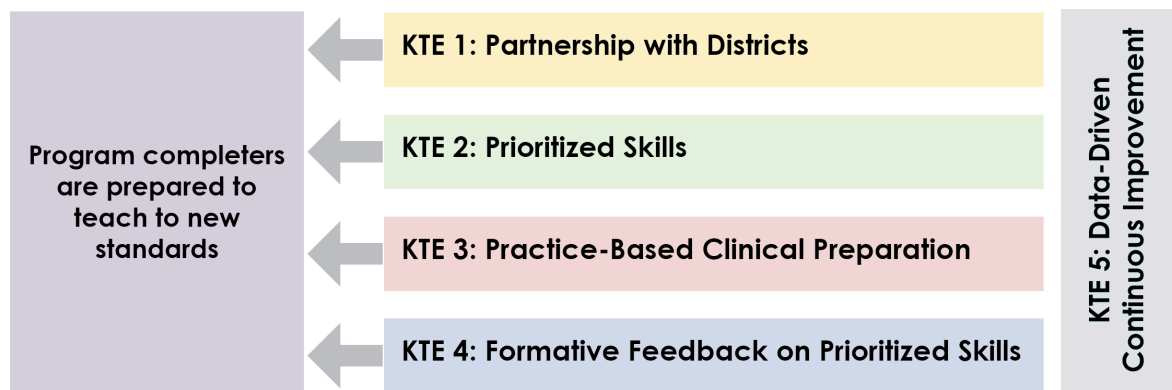
Examining process measures for TPPs that are involved in NGEI is best done in the context of understanding NGEI’s theory of action and the major processes that are involved in TPPs. NGEI seeks to strengthen the teacher preparation system in California so that new teachers enter the workforce prepared to implement the Common Core State Standards and the Next Generation Science Standards. The NGEI work, based on the most pervasive issues in TPPs, uses a theory of action to guide reform that focuses on five Key Transformation Elements (KTEs):

- Partnership with districts
- Prioritized skills
- Practice-based clinical preparation
- Formative feedback on prioritized skills
- Data-driven continuous improvement

This paper focuses on process measures for the first four KTEs, but not the fifth, because the fifth represents an approach to improving each of the first four elements (Figure 1).

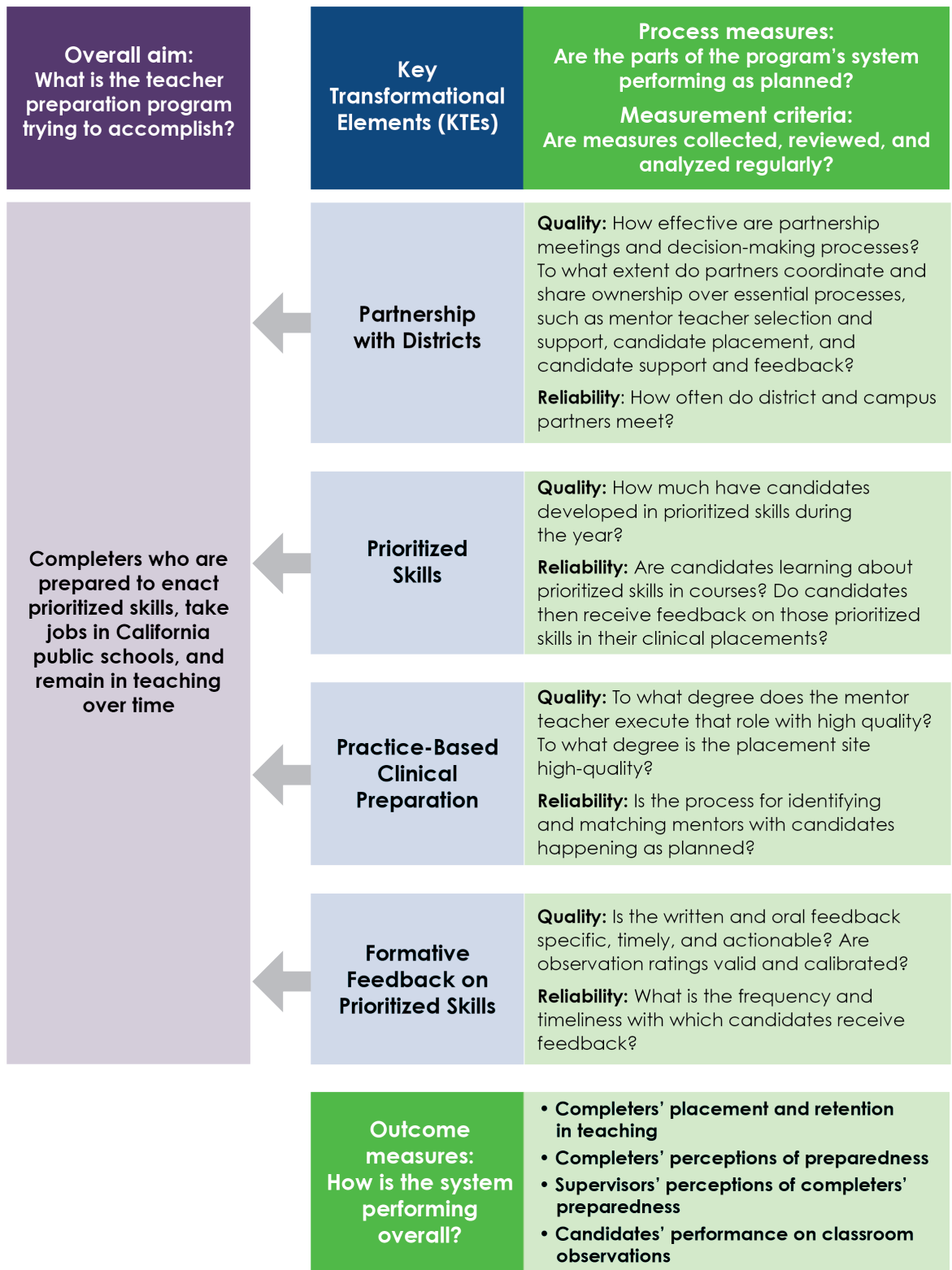
Over the past decade, a body of literature that argues for practice-based teacher education has theorized these components as essential for effective teacher preparation (Ball & Foranzi, 2009; National Center for Teacher Residencies, 2015). Programs that feature teacher residencies are often structured around key elements that are similar to the KTEs, and these programs tend to graduate teacher candidates who, on average, feel more prepared than other novice teachers in the same districts (Silva, McKie, Knechtel, Gleason, & Makowsky, 2014). There is also some evidence that participation in teacher residency programs leads to longer retention of beginning teachers (Papay, West, Fullerton, & Kane, 2011). A developing, though still inchoate, research base suggests that improvements to key systems within TPPs are factors in these outcomes. For example, providing strong clinical preparation opportunities for teacher candidates to see and practice working as teachers can support more effective teaching (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2008). In addition, research from other fields suggests that regular, high-quality feedback to candidates on their performance of key teaching practices can help them to develop these practices (Rose & Church, 1998; Scheeler, 2008).

Figure 1. Theory of Action of the New Generation of Educators Initiative



Considering each of the first four KTEs as a set of linked processes helps in determining the best ways to measure the effectiveness of TPPs in enacting each KTE. This paper outlines a starter set of process measures, examining each measure for quality and reliability (Figure 2). These suggested measures are intended for use in individual TPPs, though the measures can also be used as common measures across programs. The following sections on the KTEs examine their critical components or features, key questions about their quality and reliability, the data that might be used to track these processes over time, and examples of how some CSUs have used relevant measures.

Figure 2. System of Measures to Help Guide Teacher Preparation Programs



The following discussions of measures for each of the four KTEs also draw on three primary design principles for process measures:

- A process measure should attend to the perspectives of multiple stakeholders, including the “users” of the process.
- A process measure should capture data regularly — a rule of thumb is to collect data monthly, or more frequently, to capture indications of change over time.
- A process measure should be something for which data are relatively easy to capture, or have the potential of being easy to capture after initial development work.

Partnership With Districts

A strong partnership between a TPP and the district(s) where teacher candidates are placed for clinical practice can act as the backbone of a practice-based TPP. Candidates’ in-program experiences are situated both in the K–12 classrooms of partnering districts and in the classrooms of the teacher training institutions (which, in the case of NGEI, are CSU campuses). Thus, a coherent experience for a candidate depends on strong relationships and collaboration around all of the areas of the TPP in which the two sides of the partnership overlap. These areas include everything from mentor teacher selection and support to school site selection and agreement on priority skills. Accordingly, the strength of the partnership is important to understand and measure.

Often, a first step for partnerships between TPPs and school districts is establishing an agreement that outlines roles and responsibilities for each party, including determining key decision-makers and the resources that each partner will commit. Over time, the core work of the partnership includes selection and support of mentor teachers, sharing data and engaging in shared sense-making of data, and designing opportunities for feedback providers on both sides of the partnership to provide aligned feedback to teacher candidates, as can be done by using a common classroom observation tool (White, Torre Gibney, & Milby, 2019). In addition, ongoing communication between the district and the TPP helps to ensure that teacher candidates are receiving a positive learning experience, through the coordination of both sides of the partnership. This communication might occur through a combination of emails, phone calls, and in-person meetings (National Center for Teacher Residencies, 2015).

Reliability. TPPs need to understand whether the partnership process is enacted reliably. In order to assess the reliability of the partnership process, TPPs might consider the frequency of partnership meetings as an indicator, on the assumption that if the partners do not communicate with one another regularly, the partnership will be unable to serve as

the strong backbone that the program needs. TPPs might also consider the rate at which meetings are occurring. The ideal rate may change over the course of the partnership, from greater frequency, while TPPs and districts are developing a relationship and launching the program, to less frequency, while the program is being fine-tuned.

Quality. In measuring the quality of the partnership, the key questions to answer are: How effective are partnership meetings and decision-making processes? To what extent do partners coordinate and share ownership over essential processes, such as mentor teacher selection and support, candidate placement, and candidate support and feedback?

Short surveys of key individuals involved in the partnership — regularly asking participants from both the district and the university or college about the effectiveness of meetings, whether there is shared ownership of essential processes, and whether the decision-making process is effective and inclusive — would be one useful instrument to answer these questions.

Prioritized Skills

From the perspective of NGEI, prioritized skills are teaching skills of moderate grain size that programs, together with district partners, have selected as the most important skills for teacher candidates to learn during their preparation program. Examples of such skills include leading a group discussion, eliciting and interpreting student ideas, and building relationships with students. To ensure that candidates are provided with coherent learning opportunities throughout their program experiences, the TPPs and their district partners collaboratively identify the skills that they consider most essential for teacher candidates to learn; integrate those specific skills into coursework and the clinical experience; ensure that teacher candidates are receiving formative feedback on those skills; and assess teacher candidates on their mastery of those skills. This approach is situated in a practice-based approach to teacher preparation, which emphasizes the importance of coherent learning opportunities for candidates, organized around practices that are considered most essential to effective teaching (Ball & Foranzi, 2009).

Reliability. In evaluating whether teacher candidates are supported in the prioritized skills with reliability, TPPs may seek to answer the following questions: Are faculty, mentor teachers, and/or supervisors learning about prioritized skills and associated measurement in trainings? Are candidates learning about prioritized skills in courses? Is their learning followed by their receiving feedback on those prioritized skills in their clinical placements? To examine the reliability of training on the prioritized skills, programs could collect data on whether such training was offered, who attended, and how the attendance compared to the number expected to attend.

One approach to capturing whether teacher candidates are learning about prioritized skills in their coursework is to examine the alignment between the prioritized skills identified in the course syllabus and teacher candidates' perceptions of which prioritized skills were covered in the course (T. Flushman & S. Hegg, personal communication, April 1, 2019). In order to ensure that candidates are receiving feedback on the prioritized skills in their clinical placements, TPPs can analyze whether the mentor teachers or supervisors are giving ratings on the prioritized skills, whether they are using a classroom observation instrument, and whether they are giving feedback on these skills to the candidates.

Quality. To capture the quality of the TPP in preparing teacher candidates to learn these prioritized skills, the TPP can ask: How much have candidates developed in prioritized skills during the year? To evaluate evidence of teacher candidates developing these skills, the TPP can use a classroom observation instrument (a tool that NGEI grantees have used) that documents the teacher candidates' demonstration of skills during their clinical practice.

California State University, Bakersfield: Monitoring candidate acquisition of prioritized skills

An example from CSU Bakersfield demonstrates how one campus has used process measures to focus on the quality of its preparation of teacher candidates to learn prioritized skills. For candidates placed in the Kern Urban Teacher Residency (KUTR) program, CSU Bakersfield and partner districts have prioritized the skills outlined in the Danielson framework (Danielson, 2009) for the primary observation tool to be used in clinical placements. The Bakersfield partnership developed a detailed phase-in schedule for training teacher candidates on each prioritized skill, observing them in practice, and providing feedback. (For more details on the phase-in schedule, see White et al., 2019.) Following each observation of a student teacher by a university supervisor, mentor teacher, or district coach/instructional specialist, data are entered into an online form to capture the teacher candidate's performance on each of the skills. These data are then reviewed daily by the Continuous Improvement Lead, the KUTR coordinator, and the university supervisor. In addition, these data are brought into monthly mentor meetings, where mentors and TPP faculty discuss trends and supports for candidates. Lastly, the data are also given to district instructional specialists, to keep them informed about the progress of the teacher candidates who are in residencies in the district.

During each of these reviews, in order to know what is happening in the classroom and when to step in if necessary, CSU and district staff consider how many teachers have been rated "basic" or "unsatisfactory" on a given skill. Regarding these data, a CSU Bakersfield

team member has said, “The observation data have been instrumental to the partnership.” The data are set up so that any “basic” or “unsatisfactory” rating is highlighted in red, which allows the KUTR coordinator and the university supervisor to easily identify low ratings and provide interventions as necessary. The university supervisor explained how the partnership constantly uses the data to inform improvement: “[Almost daily,] I scroll through and look for ‘basics’ and ‘unsatisfactor[ies]’ to understand what is going on in the classroom, and would step in, especially if it is ‘unsatisfactory.’ We provide individualized support and improvement plans for residents as needed. We also look at the patterns, because if they see one ‘unsatisfactory’ score, there will be more residents with those scores. We are looking at trends and patterns. We all have access to the files. [The Continuous Improvement Lead] helps us to copy and paste the data into one giant file. We have instant access.”

The way that Bakersfield has used observation data regarding specific skills is an example of a TPP aiming to improve the performance of teacher candidates by using measures that are tied to a specific process and collected with high frequency.

Practice-Based Clinical Preparation

The third KTE focuses on the teacher candidates’ experience in K–12 classrooms, under the guidance of well-prepared mentor teachers, as part of their preparation. Among the most important components of the clinical experience is the role that mentor teachers play. A group of representatives from four national teacher preparation centers has created a shared developmental framework to improve teacher educator practice by articulating what mentors should know and do to support the candidates who are placed in their classrooms for clinical training (Beal, Comb, Dickstein-Staub, Garcia, & Salmacia, 2018). The framework specifies all of the ways that mentor teachers can effectively make explicit the work that teachers do, provide opportunities for practice, and provide targeted and continuous feedback. The measures described in this section are aimed first at ensuring that there is a reliable system for matching potential mentors with candidates, and then at ensuring that there is a system for assessing how well the mentors are able to enact the desired supports.

Reliability. The process through which teacher candidates are placed with mentor teachers is part of the larger system that supports clinical practice. Accordingly, the question that drives the inquiry into the reliability of this process focuses on matching: Is the process for identifying and matching mentors with candidates happening as planned? To answer this question, a TPP can collect data on whether, when, and how consistently key steps in the matching process occur. Reviewing such data can help the TPP ensure

that a reliable process is occurring for matching each teacher candidate with an appropriate placement.

Quality. In order to ensure that clinical placements are of high quality for teacher candidates, TPPs might seek to answer the following questions: To what degree does the mentor teacher execute that role with high quality? To what degree is the placement site a high- or low-quality match for the teacher candidate? Supervisors and teacher candidates can report on their perceptions of the quality of the mentor teachers' teaching and mentoring, to consider whether a mentor teacher is a good fit for the role. In addition, asking supervisors and teacher candidates to evaluate the quality of the placement site can provide information about the quality of the clinical placement.

California State University, Stanislaus: Gathering regular feedback about mentor teacher support

The way in which CSU Stanislaus has been working to improve the quality of teacher candidate placement provides an example of how a TPP can make use of clinical-practice process measures. Each semester, teacher candidates and supervisors complete separate end-of-term surveys (14–15 items) that provide the CSU Stanislaus TPP with feedback on the mentor teacher. Each survey includes an item, with a 5-point scale, that asks for an overall assessment of the mentor teacher and whether the respondent would recommend that teacher for future placements of teacher candidates. Items on the survey for teacher candidates are mostly parallel to those on the survey for supervisors. The final two items on the teacher candidate survey ask for an overall evaluation of the mentor teacher and about the teacher candidate's willingness to recommend that teacher for future placements.

After reviewing data from this survey, the TPP team decided that, for a placement to be considered a success, both the supervisor and the teacher candidate should rate the mentor teacher as a "4" or a "5" on the scale, as scores of "3" or lower typically indicated serious concerns. The team's goal was for at least 90 percent of candidates to evaluate mentor teachers at the top two levels, and they found that 92–95 percent of the candidates gave ratings that high. The team then looked at which of the mentor teachers were not being recommended. The team found that two mentor teachers were not recommended by candidates or by supervisors, so the team lead had a conversation with principals about those teacher placements, and Stanislaus decided not to place candidates with those teachers again.

They also found that 15–20 percent of all mentor teachers could benefit from more coaching and support. For example, the data indicated that mentor teachers were not strong at using formative data, and that teacher candidates were better at using formative

data after being matched with mentor teachers who were rated highly in using formative data.

Formative Feedback on Prioritized Skills

Feedback on teaching practice is essential for the growth and development of teacher candidates. High-quality feedback — meaning feedback that is evidence-based, frequent, specific, timely, and given by a trusted provider, and that incorporates individual learning goals — has the potential to be an important influence on teacher candidates' learning (Hannan, Russell, Takahashi, & Park, 2015; Hattie & Timperley, 2007).

Reliability. In order to measure the reliability of the feedback process, TPPs can assess the frequency and timeliness of the feedback that teacher candidates receive. With a given benchmark in mind, TPPs can determine the extent to which teacher candidates are receiving feedback at the expected frequency; whether there are undesirable delays between an observation and the feedback associated with that observation; and/or the extent to which feedback is consistently provided over a period of time.

Quality. There are a number of ways in which TPPs can measure feedback quality. The nature of either oral or written feedback can be evaluated to ensure that it is specific, timely, actionable, and tied to a set of skills prioritized by the teacher candidate and/or the TPP. This evaluation can occur through the use of a short survey that regularly captures teacher candidates' perceptions of the desired features of feedback, such as whether the feedback was specific, actionable, and/or manageable (among other qualities). Ensuring that feedback is aligned across multiple different observers ensures that teacher candidates receive consistent messages about their development of skills. This kind of norming can be measured through the monitoring of the congruence of the scores given to candidates from different observers. Several of the NGEI sites have wrestled with monitoring the quality of feedback; for example, the team at CSU San Luis Obispo has examined supervisor feedback for how it balances conveying praise versus suggesting areas of growth (Valentina & Flushman, 2017). As described in the following section, the CSU Fullerton team provides another example of examining feedback; this team looked at the relationship between the observation ratings and the evidence in written feedback.

California State University, Fullerton: Improving Quality and Reliability of Teacher Feedback

Fullerton offers an example of how to improve both the quality and the reliability of feedback received by teacher candidates. Fullerton uses the Mathematics Classroom

Observation Protocol for Practices (MCOP₂) as its primary observation protocol. All teacher candidates are observed and receive feedback through this protocol twice. For each observation, supervisors submit a performance rating on a scale of 0–3 for each of nine different student engagement indicators (of the 16 total indicators on the MCOP₂), with a written evidence statement to justify the rating. Analyzing results on these measures, Fullerton team members identified the three indicators on which teacher candidates had been receiving the lowest performance ratings and decided that they wanted to ensure that supervisors fully understood these indicators.

In spring 2018, Fullerton conducted a qualitative analysis of the evidence statements for each observation — from among more than 1,100 evidence statements for more than 100 observations — to determine whether supervisors were accurately rating teachers based on the supervisors’ written evidence. Fullerton team members coded each statement as either questionable, reasonable, or not including the indicator. They then used each of these codings to determine what proportion of evidence statements matched the rating given to teachers. For example, using the fourth indicator, “Students critically assessed mathematical strategies,” Fullerton found that more than 50 percent of supervisors were providing evidence that did not support the rating that they had provided for the observation. Consequently, Fullerton determined that, in order for supervisors to provide more accurate feedback, supervisors needed additional training on what critical assessment looks like in classrooms. Through this process, Fullerton worked to improve the quality of feedback to candidates, by ensuring alignment of evidence to ratings that teacher candidates were receiving from supervisors. Fullerton has engaged in this review of feedback quality only once to date, but regular, repeated sampling of a subset of feedback could enable use of this approach on a more frequent basis.

In addition to measuring the quality of feedback, Fullerton has also worked to measure the reliability of its feedback process, through determining the extent to which the feedback protocol was standard across its single-subject, multiple-subject, and special-education programs. In spring 2018, Fullerton issued a survey to supervisors and clinical coaches, asking them to rate the extent to which nine different steps of a feedback protocol were present, including steps such as “Start the conversation with a positive statement about what the candidate has done well and can build upon” and “Creating ‘I will’ reflection statements or targets as a way of encouraging candidates to take action on the things that have been discussed.” With 15–20 responses, Fullerton was able to identify steps that had not always been used by supervisors and/or clinical coaches. The Fullerton team has used the results of this survey to review its observation protocols in order to ensure that critical steps are taken for each observation. Through measuring the reliability of its feedback process, Fullerton has been able to work to improve the

reliability, by ensuring a standard process through which each teacher candidate can receive feedback consistently across its programs.

Limitations and Challenges

Establishing systems that can support the types of measurement approaches presented in this paper would involve some significant changes to the ways in which data are collected and processed for many TPPs. Having the data infrastructure in place to allow frequent and regular use of high-quality data to inform both support for teacher candidates and programming entails having roles and routines for collecting, cleaning, analyzing, and visualizing data that can be considered and reviewed by those who are positioned to act on the resulting learnings.

The experiences of CSU campuses involved with NGEI have revealed the challenges of having clearly defined roles to support strong and frequent data routines. For example, in the data collection phases, some project leaders have reported that their observation data are not calibrated or normed among observers, that observation data are inflated, and/or that observation data are lacking in variability. Others have reported that data analysis has often occurred too infrequently or too late to make meaningful changes. Citing inefficient data collection, management, and analysis processes, sites have found it challenging to align the timing of when data are ready for analysis with opportunities to make changes or midcourse corrections. Finally, finding time when all key staff can regularly meet to review data has been a logistical challenge for most CSU campuses involved with NGEI.

Nonetheless, many NGEI sites, including those highlighted in this paper, have taken significant steps to strengthen their data infrastructure, including establishing new roles and routines or reconfiguring existing ones. Some sites have automated elements of the data collection process, saving invaluable time. One site has been restructuring its observation data routines to allow for quarterly analysis of clinical-practice observation scores, rather than end-of-year or end-of-semester analysis. Another site has found success with a strategy for making space in packed meeting agendas by specifically designating time for “data discussion protocols.” Team leads have found that designating time for data discussion has lowered barriers to participation, helped discussions move more efficiently, and encouraged meeting participants’ interest in data, by scaffolding data discussion in an engaging and accessible way.

Conclusion

This paper highlights the need for teacher preparation programs to be focused on more than just annual outcome measures. In order to improve programs, leaders, faculty, and staff of teacher preparation programs should consider how process measures can be incorporated into their work, in order to inform ongoing continuous improvement. This paper suggests measures that could indicate both the quality of major processes and the reliability with which these processes occur as intended.

To support these approaches to measurement, it is likely that programs' data infrastructures for collecting, storing, and visualizing data on these measures will need revision to various extents. And finally, process measures and data are of little value unless they are meaningfully used to advance change. Leaders and staff of teacher preparation programs should consider the ways in which data can be incorporated into regular ongoing meetings, what context can be provided for data, who needs to be part of discussing the data, and how the data are being used to determine whether change efforts are needed and to what extent they are working.

Especially when taken together, the identification of process measures, collection of data, and use of results are critical to the ongoing improvement of teacher preparation programs.

References

- Ball, D. L., & Foran, F. M. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education, 60*(5), 497–511. DOI: 10.1177/0022487109348479
- Beal, S., Comb, M., Dickstein-Staub, S., Garcia, N., S., & Salmacia, K. (2018). *Developing the teacher educator*. Baltimore, MD: Association of Colleges for Teacher Education.
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2008). *Teacher preparation and student achievement* (NBER Working Paper, No. 14314). Cambridge, MA: National Bureau of Economic Research.
- Bryk, A. S., Gomez, L., Grunow, A., & LeMahieu, P. (2015). *Learning to improve: How America's schools can get better at getting better*. Cambridge, MA: Harvard Education Publishing.
- Danielson, C. (2009). *Implementing the framework for teaching in enhancing professional practice*. Alexandria, VA: ASCD.

Deans for Impact. (2016). *Practice with purpose: The emerging science of teacher expertise*. Austin, TX: Author.

Feuer, M. J., Floden, R. E., Chudowsky, N., & Ahn, J. (2013). *Evaluation of teacher preparation programs: Purposes, methods, and policy options*. Washington, DC: National Academy of Education. Retrieved from http://www.naeducation.org/xpedio/groups/naedsite/documents/webpage/naed_o85581.pdf

Hannan, M., Russell, J. L., Takahashi, S., & Park, S. (2015). Using improvement science to better support beginning teachers. *Journal of Teacher Education*, 66(5), 494–508.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. DOI: 10.3102/003465430298487

IHI Multimedia Team. (2015). *Like magic? (“Every system is perfectly designed...”)*. Boston, MA: Institute for Healthcare Improvement. Retrieved from <http://www.ihl.org/communities/blogs/origin-of-every-system-is-perfectly-designed-quote>

National Center for Teacher Residencies. (2015). *Clinically oriented teacher preparation*. Chicago, IL: Author. Retrieved from <https://nctrresidencies.org/wp-content/uploads/2015/07/NCTR-COTP-Final-Single-Pgs.pdf>

Papay, J. P., West, M. R., Fullerton, J., & Kane, T. J. (2011). *Does practice-based teacher preparation increase student achievement? Early evidence from the Boston Teacher Residency* (NBER Working Paper No. 17646). Cambridge, MA: National Bureau of Economic Research.

Provost, L. P., & Murray, S. K. (2011). *The health care data guide: Learning from data for improvement*. San Francisco, CA: Jossey-Bass.

Rose, D. J., & Church, J. (1998). Learning to teach: The acquisition and maintenance of teaching skills. *Journal of Behavioral Education*, 8(1), 5–35. Retrieved from <http://www.jstor.org/stable/41824208>

Scheeler, M. C. (2008). Generalizing effective teaching skills: The missing link in teacher preparation. *Journal of Behavioral Education*, 17(2), 145–159. Retrieved from <http://www.jstor.org/stable/41824430>

Silva, T., McKie, A., Knechtel, V., Gleason, P., & Makowsky, L. (2014). *Teaching residency programs: A multisite look at a new model to prepare teachers for high-need schools* (NCEE 2015-4002). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Southern Regional Education Board Teacher Preparation Commission. (2017). *Teacher preparation data systems: State policy and recommendations*. Atlanta, GA: Author.

Retrieved from: https://www.sreb.org/sites/main/files/file-attachments/teacherprepcommission_datareport_2017.pdf

Valentina, V., & Flushman, T. R. (2017). Emphasis of university supervisor feedback to teacher candidates. *Journal of Student Research*, 6(2), 45-55.

White, M. E., Torre Gibney, D., & Milby, A. (2019). *Developing systems for high-quality feedback to teacher candidates: Lessons learned from 11 California State University teacher preparation programs*. San Francisco, CA: WestEd.