Abstract

Data from program assessment in higher education are often used for accreditation purposes and are less focused on decision-making and program improvement. This article illustrates how data-informed decisions were made in a teacher education program. It details how a framework of assessment, Plan Do Study Act, was used to identify areas in need of attention and how the program made feasible incremental changes to its curriculum and assessment process over a three-year period to improve students' learning.

AUTHORS

Ya-Chih Chang, Ph.D.
California State University, Los Angeles

Holly M. Menzies, Ph.D.
California State University, Los Angeles

Data-Informed Decision-Making in Higher Education: Lessons from a Teacher Education Program

As knowledge and skill in assessment practice matures, universities are fostering processes better aligned for use at the program and instructor level. There is a long-standing tension between institutional assessment for improvement and its use for accreditation purposes. While these purposes should be the same, they are more often seen as a "contradiction" (Ewell, 2009, p. 7) because of the accountability mandates inherent in accreditation requirements. However, there has been a shift toward what Ewell (2009) characterized as the "improvement paradigm" and away from the "accountability paradigm."

The focus of an improvement paradigm is for faculty to identify and collect evidence of student work to examine whether students are mastering course and programmatic outcomes and determine whether changes are needed to improve student learning. This contrasts with an accountability stance where the purpose is to signal to an external audience the worthiness of the institution, typically through standardized measures or institutional-level metrics (Ewell, 2008). Blaich and Wise (2010) were among the first to assert that assessments do not necessarily lead to improved learning. They emphasized the importance of assessment but also noted that assessment processes in higher education are frequently more political than data-driven; thus, assessment leaders must understand the social, political, historical, and budgetary context of the institutions to make pragmatic choices about which assessments to administer. This may mean collecting data that are most interesting to faculty, even if it does not directly result in student learning improvement.

Other researchers have proposed assessment models that address the importance of student learning outcomes, including "closing the loop" (Banta & Blaich, 2011), learning improvement (Stitt-Berg et al., 2018), Program Learning Assessment, Intervention, Re-assessment (PLAIR; Fulcher et al., 2014), and Plan, Do, Study, Act

CORRESPONDENCE

Email
ychang27@calstatela.edu

(PDSA; Moen, 2009). These different methods are suited for use at the local level to improve student outcomes. Each foregrounds a practical approach to collecting evidence, making changes, and evaluating the impact of those changes, although each method employs slightly different tactics. Additionally, they all emphasize direct assessment of student learning and a formative approach to instructional and program improvement. For example, the PLAIR model focuses on change and intervention, instead of only on the assessment methodology (Fulcher et al., 2014). Implementing this model may take a few years as programs must assess, identify the area that needs improvement, develop and implement the appropriate intervention, and then reassess to determine whether student learning improved. The PDSA framework is a similar model that also uses a "closing the loop" approach (Moen, 2009). However, it operates on a faster timeline than the PLAIR model. For example, the PDSA allows programs to make changes to the program before the first cohort of students graduates, depending on what is assessed.

A major benefit of a PDSA cycle is the focus on rapid and iterative change. If a change does not result in the desired outcome, another change can be tested.

This case study reports program-level efforts to use a PDSA framework to improve student learning in an Early Childhood Special Education (ECSE) program where students earn a state-issued credential to teach in an early childhood setting and work with young children with disabilities. As faculty adopted formative processes to make programmatic decisions, they faced measurement issues and implementation challenges in making meaningful changes to courses. This paper details the assessment process and decisions made based on program-level information collected about student learning.

Plan, Do, Study, Act

The PDSA cycle was originally used in business to emphasize continuous improvement and was popularized by Charles Deming in the mid-twentieth century (Moen, 2009). It consists of four elements for making iterative changes as part of an ongoing information collection and analysis cycle. The first step is to *Plan* a change that will improve an outcome. The change can be based on formal quantitative data collection, but it can also be from qualitative data collected from faculty or students. While it is important that data have integrity, it is also important not to wait for perfect information as the data or tools used to collect it will invariably have some flaws (Berwick, 1996). The second step, *Do*, is to put the change into place. Once information or data have been collected on the results of the change, you Study to determine whether a positive change has occurred in the targeted outcome. *Act* is deciding whether to permanently adopt the change or try a different change based on the analysis. It can also mean the continuation of data collection for progress monitoring.

A major benefit of a PDSA cycle is the focus on rapid and iterative change. If a change does not result in the desired outcome, another change can be tested. The cyclical nature of the model capitalizes on the use of ongoing information collection and analysis for monitoring improvement. The PDSA model can help programs plan and implement beneficial changes, and the process itself is straightforward.

Context and Process

Institutional Context

The university is located in a dense urban area in the West. It serves a diverse population of primarily first-generation college students. After the university's 2010 accreditation by the Accrediting Commission for Schools, Western Association of Schools and Colleges (WASC), additional resources were directed to assessment practices. Assessment efforts and resources included the establishment of a university-wide committee to promote the use of assessment, a variety of training offered to faculty (which is ongoing), and the formation of an assessment committee within each college. A member of each college committee sits on the university committee to facilitate assessment work across the institution.

The College of Education assessment committee is comprised of two representatives from each of the three departments in the college, the chair of each department, and the associate dean of the college, who serves as the committee chair. The committee predates the 2010 WASC visit as there are over 25 accredited programs in the college, and assessment work was integral to maintaining accreditation status. While compliance concerns were

undoubtedly a driver of instituting assessment practices, programs were encouraged to develop systems tailored to their specific needs. Also, programs had different assessment requirements depending on their accreditation body, so flexibility in assessment practices was crucial.

A major focus of the committee is to guide programs in using data-informed decision-making to improve teaching and learning and address accreditation requirements. Committee members attended assessment-related trainings and read and discussed a series of books and articles on assessment. Additionally, the state university system provided the committee and the associate dean with coaching in improvement science to enhance their ability to improve student outcomes. The PDSA model was one method the committee had explored together. Programs had encountered various challenges using data to make curricular or instructional changes as faculty tended to focus on the quality and quantity of the data instead of "closing the loop." While validity is an important consideration, it had become a barrier to substantive improvement. PDSA was a low-stakes method for focusing on actionable and formative information. Accreditation requirements are typically focused on global indicators such as pass rates on state-mandated exams and the percentage of students successfully completing fieldwork, but these did not provide the more granular information needed for program improvement. However, accreditation for teacher credentialing did require programs to demonstrate how they collected data to make programmatic decisions.

Department Assessment Processes

The Department of Special Education and Counseling includes several special education credential options (e.g., ECSE, visual impairments, mild/moderate support needs, and extensive support needs). Each of the special education program options evaluated candidates during their fieldwork practicum using similar formative and summative measures to allow comparison across options while still evaluating competencies specific to the credential area. The program coordinator for each option facilitated data collection and then aggregated and analyzed the data. Results from each program were discussed annually in a department meeting. Next, we describe how one of these options, ECSE, used the PDSA cycle to make data-informed changes for program improvement.

Early Childhood Special Education Program

Overview of the Program

The ECSE Program is a two-year program that prepares teacher candidates to serve young children (age 0-5) who are at risk or with a disability. The program follows an intentional sequence of coursework to first introduce candidates to foundational knowledge in disability, characteristics of children with special needs, special education law, first and second language acquisition in the context of cognitive development, social emotional development, and classroom management and positive behavior support. Subsequent coursework uses this grounding as context for developing knowledge for assessing, planning, and providing learning opportunities for infants, toddlers, and preschoolers. Candidates demonstrate knowledge and teaching competencies in a final student teaching experience during their last term in the program where they are placed in an early childhood setting with young children (ages 0-5) with and without disabilities.

Programs had encountered various challenges using data to make curricular or instructional changes as faculty tended to focus on the quality and quantity of the data instead of "closing the loop."

Student Teaching Fieldwork Course

Approximately 25-30 candidates enroll in the student teaching fieldwork course each year. Most fieldwork placements are in high-needs schools (e.g., low SES, Title I). At the end of the course, candidates are expected to demonstrate teaching competencies in the areas of Assessment, Curriculum, Managing the Teaching and Learning Environment, and Collaboration and Professionalism to be recommended for a credential to the state credentialing authority.

University Supervisors

All university clinical supervisors hold an ECSE credential or have experience in ECSE. The program coordinator meets individually with newly hired university supervisors to review the fieldwork requirements and explain how to administer the assessment measures. There is a cadre of experienced supervisors who are assigned to candidates each semester and occasionally a new supervisor is hired. Most clinical supervisors are adjunct faculty, but also include tenure line faculty. University supervisors meet with each candidate a minimum of six times over the course of the term. When conducting an observation, the university supervisor completes a formative assessment measure to evaluate and provide feedback to the candidate on their teaching. This measure is completed electronically, making it simple to collect, aggregate, and analyze the data at the end of each term. A single summative measure is completed at the end of the term to indicate the candidate's proficiency level for each of the competency domains. Candidates also receive structured feedback from their cooperating teacher, or supervising administrator, if they are interns.

A new observation form was developed that included the same global domains as the summative measure but comprised discrete items that could be rated to provide more specificity about candidate performance.

Measures

The Early Childhood Special Education (ECSE) program uses both summative and formative assessment tools to evaluate the program, provide feedback, and determine whether candidates meet program competencies.

Summative measure

The summative rubric assesses skills in four competency domains: Assessment, Curriculum, Managing the Teaching and Learning Environment, and Collaboration and Professionalism. A five-point scale ranging from 1 (Preliminary) to 5 (Mentor Level) is used to describe performance. Each level includes a detailed narrative description of performance. Expected performance at the end of the semester is a score of 3 in each domain with a total summed score of 12. Mentor level is included because many candidates are interns and have considerable teaching experience by the time they complete their program.

Formative measure

Originally, supervisors provided only written feedback after each observation; however, this was cumbersome to aggregate and report and challenging to track over time. A new observation form was developed that included the same global domains as the summative measure but included discrete items that could be rated to provide more specificity about candidate performance. For example, in the assessment domain, one item is "Provides timely and high-quality feedback to students about lesson content/material." Supervisors rate performance on each item using a scale ranging from 1 (does not meet standard) to 4 (exceeds standard). Items can also be rated as not applicable (N/A). The formative measure includes areas for narrative comments and collects demographic information, such as whether the candidate is an intern or a traditional student teacher. This allows the program to analyze data with more precision (See Appendix 1 for Sample Items from Formative Assessment).

Data Analysis

We report retrospectively on the program improvement process and describe the steps taken for program improvement. This endeavor was not originally conceived as a research study, so an a priori data analytic plan was not created. However, the study details the processes in how we implemented a new model to inform programmatic changes.

Each term, the formative measure data uploaded by an individual university supervisor was retrieved and stored on the department's SharePoint site. The data from the summative measure were entered into an Excel spreadsheet as the form was completed on a paper copy. It, too, was stored on SharePoint. At the end of each academic year, the data are aggregated and examined by the tenure line faculty in the program.

Data described in this case study were collected from five different supervisors who observed two to three candidates each term (approximately 25-30 candidates per academic year). Supervisors completed the formative measure at each observation and the summative

measure at the end of the term. Descriptives, including means and standard deviations, were calculated for approximately 100 observations yearly over the course of three years.

The ECSE PDSA Cycle

Below we describe the three-year iterative process of using fieldwork assessment data to initiate and evaluate programmatic changes (See Table 1). These included the development of a new clinical course, modification of assignments and readings in existing courses, and refresher training for university supervisors.

Table 1. PDSA Cycle

	Year 1	**Year 2**	**Year 3**
Plan	Assess how well the program prepared candidates in working with young children with and without disabilities in naturalistic classroom settings.	Assess how well the program prepared candidates in working with young children with and without disabilities, specifically in two areas: 1. Candidates' competency in assessments 2. Candidates' understanding and effective use of technology in early childhood classrooms	Assess how well the program prepared candidates in working with young children with and without disabilities. 1. Continue to monitor candidates' competency in assessments 2. Establish university supervisors' reliability and consistency in scoring students' use of technology in classrooms
Do	Data collection using both formative and summative assessments during final fieldwork.	Data collection using both formative and summative assessments during final fieldwork.	Data collection using both formative and summative assessments during final fieldwork.
Study	1. Assessments had relatively low scores compared to the three other domains that were evaluated during fieldwork. 2. University supervisors rated "N/A" in student teachers' use of technology.	1. Candidates demonstrated an increase in their competencies in the domain of Assessments. 2. University supervisors continued to rate "N/A" in student teachers' use of technology.	1. Candidates continued to demonstrate competencies in the domain of Assessments. 2. There was a significant decrease in "N/A" ratings on student teachers' use of technology.
Act	Modify two courses in the program to increase candidates' competency in assessment practices and effective use of technology.	Redesign supervision training to improve supervisors' administration of the formative assessment.	No modifications were made. Continue to monitor candidates' competency in all domains.

Year 1

Plan

In Year 1, the faculty decided to triangulate the results of both the formative and summative measures to determine how well the program prepared candidates to work with young children with and without disabilities in naturalistic classroom settings.

Do

Each term, university supervisors used the newly created formative measure to provide feedback to candidates during each in-person observation. They also used the summative assessment rubric to determine whether candidates met all teaching competencies at the end of the term. The data for the formative measure were entered electronically during each visit, and the summative measure was completed by hand using a paper form and later entered into an Excel database. At the end of the academic year, the program coordinator aggregated and analyzed both datasets.

Study

Summative data indicated candidates scored relatively low in the domain of Assessment. This domain included: (a) the selection and use of multiple, appropriate, formal and informal non-biased assessment tools with consideration of cultural, linguistic, and ability status across developmental and educational domains; (b) monitoring of student's progress regularly with data-based, anecdotal, and authentic input from all team members; and (c) appropriate adaptation of student programs in response to regular assessment of progress across developmental and academic domains. Average scores ranged between 3.7-4.0 (Advanced/ Independent Level) for all domains except in *Assessment*, where they were at approximately 3.2 (Proficient/ Beginning Teacher - Advanced/ Independent Level), a noticeable difference in contrast to average scores in the other domains.

The results helped faculty identify two areas for program improvement for the following year: 1) Competency in assessment practices and 2) Effective use of technology in early childhood classrooms.

The formative measure indicated lower scores in using formative data to develop lesson plans aligned to the Preschool Learning Foundations and providing specific feedback to children. It was also evident that university supervisors frequently used the "N/A" descriptor on items related to technology in the classrooms. These items were "Effectively uses varying levels of technology (low tech/high tech) to meet student needs specific to classroom management and whole class participation" and "Integrates technology (low tech/high tech) to enhance student engagement and address learner needs specifically to lesson/content learning." This was a concern as it signaled that technology was either not being adequately used by candidates or supervisors were having difficulty distinguishing what constituted technology use.

Act

The results helped faculty identify two areas for program improvement for the following year: 1) Competency in assessment practices and 2) Effective use of technology in early childhood classrooms. The faculty in the ECSE program decided to make modifications to two courses in the program to increase candidates' competency in the identified areas of concern.

Year 2

Plan

In Year 2, the program continued to collect assessment data on how well the program prepared candidates. In addition, they made modifications to two courses as decided at the end of Year 1. The first modification was made to the assessment course which focuses on understanding how different assessments are used in early childhood, including standardized and formative assessments. The final assignment for the course is to write an assessment report of a young child (ages 0-5) with disabilities and include three clear goals based on data collected during the term. Candidates are required to use three different types of assessments (e.g., parent interview, classroom observation, standardized assessment) and include progress monitoring data in the appendix of the report. This assignment was modified in Year 2 to include a reflection on the type and usefulness of the data collected with the goal of making candidates more intentional about their assessment practice.

The second change was the addition of a new clinical course. Its purpose was to bridge knowledge and clinical practice by providing candidates with additional hands-on experience working with children with and without disabilities in a diverse inclusive community enrichment program. Candidates were to enroll in it during the first year of the program

and it was to be taken in conjunction with a methods course that included topics such as classroom management, routine building, early language and literacy, play, and technology use in early childhood settings. To address the issues identified about assessment and technology use, candidates were asked to write developmentally-appropriate lesson plans and implement evidence-based strategies under the supervision and coaching of course instructors. Candidates were required to assess children's understanding of the lessons and monitor children's progress throughout the term. Candidates were also expected to use technology (e.g., short videos) in their lessons each week. Readings on technology use in early childhood settings were assigned to increase candidates' understanding and effective use of technology and provide candidates with different examples of technology use in early childhood settings.

Do

University supervisors continued to use formative and summative assessments to determine whether candidates met teaching competencies, the new course was offered, and the proposed curricular changes were made.

Study

Data from the summative assessment showed that on average, the candidates' scores in Assessment were higher. Average scores were now in the 4.4 (Advanced/ Independent Level), comparable to those in the other three domains. This suggested that the course modifications increased candidates' understanding and practice in using assessments for planning instruction.

However, there was no change in average scores for the technology items on the formative measure. Despite the additional course readings and practice of technology use in the Early Intervention Lab course, university supervisors continued to frequently rate "N/A" in candidates' technology use.

Act

Candidates demonstrated an increase in their scores in the domain of Assessment; therefore, the course changes were made permanent. Candidates' use of technology remained an area of concern. It was hypothesized that candidates may have been effectively using technology, especially after the curricular modifications from the previous year, but the problem may be one of measurement. Therefore, the faculty decided to refresh university supervisors' knowledge of how to score candidates' use of technology to improve their administration of the formative measure.

Year 3

Plan

In Year 3, the program faculty continued to collect and monitor assessment data. They redesigned the supervisor training process to improve reliability and consistency in scoring candidates' use of technology in the classrooms.

Data from the summative assessment showed that on average, the candidates' scores in Assessment were higher. Average scores were now in the 4.4 (Advanced/ Independent Level), comparable to those in the other three domains.

Do

University supervisors were invited for refresher training on using the formative assessment rubric for final student teaching. The program coordinator led the training which was approximately two hours long. University supervisors were provided with a small stipend to attend the training. Prior to the training date, supervisors were sent two 15-minute videos of an ECSE lesson. Each supervisor was asked to evaluate the two videos using the formative measure and provide their ratings.

The goal was to re-calibrate scoring across all university supervisors in the program. At the first meeting, program coordinators emphasized the items where there were substantial differences in the use of the N/A category. Supervisors were reminded that any use of technology, both low tech (e.g., individualized communication boards) or high tech (e.g., use of iPad for short videos), should be scored. Additionally, specific descriptions for each rating

level (e.g., 1, 2, 3, and 4) were added for each item based on the discussion at the meeting. See Appendix for examples of descriptors added. Using the revised descriptions, the group re-scored the first two videos and came to a consensus on their ratings.

The PDSA cycle was a valuable process for making meaningful and consequential changes for program improvement.

After the meeting, the university supervisors were provided with another set of two training videos and were again asked to rate the videos using the revised rubric. The program coordinator reviewed the ratings and found the ratings to be more reliable across university supervisors.

Study

At the end of Year 3, candidates' average scores on the summative assessment in Assessment continued to be in the range of 4 (Advanced/ Independent Level), as did the other three domains. There was a significant decrease in the rating of "N/A" on the formative assessment for technology use. Instead, students were rated highly for using appropriate technology to advance the quality of their lessons.

Act

No curricular or program modifications were planned for the following year because average scores had increased to an acceptable level. Data collection and analyses were retained to examine program quality on an ongoing basis.

Discussion

Using the PDSA model, program faculty were able to identify areas in need of change and make improvements over the course of three years. This process was useful for accreditation purposes because it documented how programmatic changes were data-informed, but more importantly, it enabled incremental and feasible changes that improved the *quality* of the program.

In the first two years, course improvements were made that addressed program competencies in using assessments, including collecting and analyzing data on young children's skills in various developmental domains. Assignments and courses were modified or added to provide more contextual opportunities for candidates to practice these specific skills. For example, an early intervention lab was instituted that required candidates to collect data about their students and monitor progress over the course of the term. Student teachers identified areas of need, developed lessons, and implemented them with embedded learning opportunities to meet their young students' individual needs. The early intervention lab was offered during the first year of the program offering candidates the opportunity to practice assessment skills before being evaluated for competency in the fieldwork practicum. In the third cycle, a measurement issue was identified. After two years of curriculum modifications to address the use of technology in early childhood classrooms, university supervisors were still frequently using the N/A category instead of rating candidates' level of proficiency on the technology items. The program decided to retrain university supervisors in identifying and evaluating candidates' use of technology. The PDSA process made it possible to bridge the divide that sometimes occurs between content learned in coursework and evaluation of its application in practice. In this case, once university supervisors received additional training on the topic of technology aligned to the coursework and understood what to look for, they were far less likely to use the N/A category.

The PDSA cycle was a valuable process for making meaningful and consequential changes for program improvement. It can be challenging to know where to start with assessment, especially when faced with the task of making changes to a program that has several courses, a variety of fieldwork experiences, and many instructors. However, using the PDSA cycle made the endeavor both manageable and productive.

References

Banta, T. W., & Blaich, C. (2011). Closing the assessment loop. *Change: The Magazine of Higher Learning*, *43*(1), 22-27. doi: 10.1080/00091383.2011.538642

Berwick, D. M. (1996). A primer on leading the improvement of systems. *BMJ*, *312*(7031), 619-622.

Blaich, C. F., & Wise, K. S. (2010). Moving from assessment to institutional improvement. *New Directions for Institutional Research*, *2010*(S2), 67-78.

Ewell, P. T. (2008). *Assessment and accountability in America today: Background and context*. In V. M. H. Borden & G. Pike (Eds.), Assessing and accounting for student learning: Beyond the Spellings Commission (New Directions for Institutional Research, Assessment Supplement 2007, pp. 7-18). Jossey-Bass.

Ewell, P. T. (2009, November). *Assessment, accountability, and improvement: Revisiting the tension*. (Occasional Paper No. 1). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Fulcher, K. H., Good, M. R., Coleman, C. M., & Smith, K. L. (2014, December). *A simple model for learning improvement*: *Weigh pig*, *feed pig*, *weigh pig*. (Occasional Paper No. 23). Urbana, Il: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Moen, R. (2009). Foundation and history of the PDSA cycle. https://deming.org/wp-content/uploads/2020/06/PDSA_History_Ron_Moen.pdf

Stitt-Bergh, M., Kinzie, J., & Fulcher, K. (2018). Refining an approach to assessment for learning improvement. *Research & Practice in Assessment*, *13*, 27-33.

Appendix

Appendix 1. Sample items from the formative assessment

1=does not meet standard 2=approaching standard 3=meets standard 4=exceeds standard NA=no lesson plan

	1	2	3	4	NA
Integrates technology (low tech/ high tech) to enhance student engagement and address learner needs, ***specifically to lesson/ content learning.*** **should have conversation with student candidate if they are not using technology to enhance lesson* 1= no use of technology, but should have 2= have technology but did not use appropriately 3= have technology AND used it appropriately 4= Appropriate use of technology that advanced the quality of the lesson and was accessible to all students	( )	( )	( )	( )	( )
Provides timely and high quality feedback to students about ***lesson/content material.*** 1= None 2= Responds to student 3= Responds AND embeds strategies 4= Consistently responding and embedding opportunities throughout activities AND provides additional content information	( )	( )	( )	( )	( )