

Executive Summary
Case Studies of Butler University Teacher Education Graduates
April 2018
Dr. Deborah Corpus & Dr. Brooke Kandel-Cisco

Faculty in Butler University's College of Education, under our leadership, planned and implemented a research project using case study methodology to explore two questions: 1) Do our program completers contribute to expected levels of student learning growth, and 2) Can we document, in the words of CAEP Standard 4, "through structured and validated observation instruments and student surveys, that completers effectively apply the professional knowledge, skills, and dispositions that the preparation experiences were designed to achieve?"

Background Information

Before documenting the impact of our graduates on K-12 student learning growth, we sought to collaboratively identify and describe the professional knowledge, skills, and dispositions that our educator preparation program (EPP) was designed to achieve. While our EPP is grounded in national standards including INTASC and Specialized Professional Associations, we also wanted to capture the unique components of our preparation program that shape the teaching practices of our graduates.

Identifying Primary Traits

Those familiar with forms of holistic and analytic scoring of student writing samples will be familiar with the introduction of primary trait scoring by Lloyd-Jones (1977). We spent a full year working with a variety of stakeholders to identify the primary traits of our EPP with the intention of building those traits into the data collection strategies and scoring guide for our CAEP Standard 4 case studies. We conducted focus groups with current undergraduate and graduate students, practicing teachers, school counselors, and principals who were alumni of our educator preparation programs, and teachers at our K-8 laboratory school. We asked focus group participants a variety of questions intended to probe their experiences in our EPP and to excavate the impact of those experiences on their teaching practices and on their K-12 students. Focus group participants named and described an extensive list of knowledge, skills, and dispositions that they believed were connected to or were outcomes of their educator preparation experience. We aggregated focus group data to develop the primary traits of our EPP.

These primary traits were validated through ongoing work with the entire College of Education faculty and staff. We also conducted analysis of College of Education documents to look for parallels between and contradictions with the primary traits articulated by the focus groups. This adaptation of primary trait scoring from the field of writing assessments allowed us to identify the those primary traits and build draft rubrics to focus our attention on those aspects of our program.

Methods

The state of Indiana does not collect Value Added Measures (VAM) and does not release student test data connected to specific teachers. In order to gather data about student achievement, we used a case study method to describe and explain the impact of the EPP on our graduates’ teaching practice and to explore the potential impact of our graduates’ practice on K-12 student learning. Our multiple case design, wherein each teacher was a case, allowed us to study the practices of our graduates within their unique teaching contexts. We sought to answer two research questions examining data across a variety of contexts and through multiple data sources to answer these questions: 1) Do our program completers contribute to expected levels of student learning growth, and 2) Can we document, in the words of CAEP Standard 4, “through structured and validated observation instruments and student surveys, that completers effectively apply the professional knowledge, skills, and dispositions that the preparation experiences were designed to achieve?”

Participants

In an effort to ensure our data reflected the range of content areas and developmental levels in which we license teachers, we selected and invited potential participants based on proximity, content area, and developmental level. After two years of data collection, our purposive sample included six graduates who were in their second or third years of teaching.

Participants in Year 1 and Year 2			
Year	Teacher	Grade and content area observed	School
2016-17	1	Grades K-1;	K-8 Laboratory school
2016-17	2	Grades 6-8; Language Arts	K-8 Laboratory school
2016-17	3	High school; Physics	Urban public school with COE partnership
2017-18	4	Grade 3; Mathematics	K-8 Laboratory school
2017-18	5	High school; Algebra	Urban public school with COE partnership
2017-18	6	Grade 6; Social Studies	Suburban public 6-8 school

Data Sources & Triangulation

We used multiple data sources to describe the impact of our EPP on our graduates’ teaching practices and to explore the potential impact of our graduates’ practice on K-12 student learning. The process for

collecting data on each case included a pre-observation interview, a classroom observation, and a post-observation interview. We asked participants to share a unit of study, however they defined that unit, that they were expected by their school to teach. They were asked to plan, teach, and evaluate using whatever forms and structures their school required or ones they normally used. At least two researchers interviewed the teachers, observed the teachers teaching, and conducted a post-teaching interview using previously developed data collection protocols. Using two researchers allowed a more thorough collection of information. The multiple data sources are the following:

Pre-observation interview. In the semi-structured pre-observation interview, the two professors scheduled to observe the lesson asked the teacher to describe the unit of study and explain how the lesson we would observe supported the goals of the unit. Interview questions invited participants to name the key knowledge and skills taught in the lesson, describe intended student learning outcomes, and identify specific needs of individual students or groups of students. Each researcher recorded notes of the teacher's responses. The notes of each researcher were included in the data set. We also asked participants to describe their plans for assessment in the lesson and their expectations for student growth as documented in the assessment plan. The teachers provided us a lesson plan or other lesson documentation, and we added those documents to our data.

Direct classroom observation. Within one week of the pre-observation interview, at least two observers conducted a direct classroom observation using an observation protocol based on the primary traits of our EPP. The observers recorded the teacher's behaviors, language choices, and teaching practices as well as student behaviors and responses. The goal of the direct classroom observations was to develop thick descriptions of evidence of student-learning growth and application of the knowledge, skills, and dispositions identified as primary traits and foundational to our preparation program. Each professor transcribed their notes into a frame organized by Standards 4.1 and 4.2 and their identified subcomponents.

Post-observation interviews. The two researchers who observed the teacher and students met again with the teacher and conducted a post-observation interview using a standard set of questions focusing on Standards 4.1 and 4.2. The notes were organized by standard and subcomponents. Both sets of notes from the researchers became part of the data set.

Physical artifacts. Lesson plans and class- and student-level data relevant to the observation and interviews were collected by the professors who conducted the interviews and observations. These artifacts became part of the data set.

The pre-/post- interviews and observations by two researchers resulted in six sets of data in addition to the physical artifacts. All of this data was organized into a frame by Standard 4.1 and 4.2 and their identified subcomponents so that scorers could easily access all data.

Procedures

We led the CAEP Standard 4 efforts and were aided by a team of faculty volunteers. After we had planned the framework for a case study approach to answer the questions posed by CAEP Standard 4 and the examination of the attributes that make the Butler University teacher preparation program

unique, we developed the Standard 4 Program Impact Case Study Rubric, collected the data, trained the observers and scorers, and began analyzing the data.

Rubric Development

We drafted a Standard 4 Program Impact Case Study Rubric to use in evaluating the five sub-components identified by us as part of Standard 4.1 and the three sub-components identified as part of Standard 4.2 using the information from the multiple data sources. The rating system is the four-point scale used by the College of Education on all its assessments: 1: Emergent, 2: Basic, 3: Competent, 4: Proficient. The system also contains a category of “Not Observed.” The descriptions of each score point were extensively revised as it was reviewed by different groups of stakeholders: College of Education faculty, graduate students in our masters program, faculty in our Lab schools. The rubric was tested in 2017 using the data from our first year’s case studies, and some language was clarified based on the recommendation of the evaluators. The rubric was sent to CAEP for early instrument feedback and the descriptive language clarified further in March of 2018 based on recommendations by the CAEP reviewer.

The data from the interviews, the observations, and the artifacts was then analyzed using the Standard 4 Program Impact Case Study Rubric designed specifically to examine the two questions posed: Is there evidence of student-learning growth? (CAEP 4.1) Do program completers apply the knowledge, skills, and dispositions that their preparation was designed to achieve? (CAEP 4.2)

Initial Training and Procedural Revisions

Initial training in Spring 2017 for both observation and for use of the Program Impact Case Study Rubric was based on videos of teachers and sample documentation from edTPA submissions by our student teachers. The observers and readers met together as they used the data collection forms and then used the Program Impact Case Study Rubric. The observers and scorers met together to develop a shared understanding of the process and the use of the data. They compared notes of what was considered important in their data collection. They discussed discrepancy in case study scores for the practice data to clarify language on the rubric and interpretation of data.

We then began the process of interviewing, observing, conducting a post-interview, and collecting artifacts for the first group of teachers. We expected to make refinements to the process, the data collection documents, and the scoring rubric based on this first experience. We took notes of changes that would expedite the process or provide us with more helpful information.

These notes from our first experience in collecting data and using the Program Impact Case Study Rubric resulted in some refinements to the organization of our data. Participants in the observations requested sample indicators for each subcomponent be embedded in the classroom observation form to help focus attention. Participants in the scoring sessions requested that data be organized by subcomponent of each standard to streamline finding the data and reading it during the scoring sessions. This resulted in our developing the Frame for Compiled CAEP Standard 4 Evidence which we used in Year 2 to organize the collected data for our scorers.

Scoring and Analysis

Scorers in Year 2 were trained and calibrated to the Standard 4 Program Impact Case Study Rubric and the Framework for Compiled CAEP Standard 4 Evidence using the data from a Year 1 case study. The scorers read through the data for the first subcomponent and then discussed the score. The scorers then worked independently, subcomponent by subcomponent scoring, comparing scores, and discussing until they reached an acceptable level of inter-rater reliability.

Findings

Findings from the six cases in Year 1 and Year 2 are summarized in the table below. The researchers examined the data across participants, while also attending to patterns within elementary and middle/secondary developmental levels.

Components identified by Butler Stakeholders	CAEP Standard 4 Program Impact Case Study Rubric based on a 4-point scale with 4 being highest	Key findings
4.1.1. Evidence of Student Learning Growth: Identification of knowledge and skills to be taught in a unit.	Year 1: 3, 4, 4 Year 2: 2, 4, 3	Middle and secondary graduates in our sample identified and connected the knowledge and skills to be taught to the lesson. Teachers at our Lab School, both elementary and secondary, noted school-based standards. Teachers in other contexts noted state standards.
4.1.2 Match between the knowledge and skills taught and evaluations.	Year 1: 3, 4, 3 Year 2: 2, 2, 2	Results show variability across the two years of data collection. We believe the variability has more to do with changes made to the language of the rubric than to differences in teacher practice.
4.1.3 Evidence of formative assessment throughout the unit.	Year 1: 4, 4, 4 Year 2: 4, 3, 3	Graduates in our sample used formative assessments in their lessons. 4 out of 6 graduates used multiple types of formative assessments.
4.1.4 Teacher provides evidence of student learning.	Year 1: 4, 4, 3 Year 2: 4, 4, 3	Most graduates (4/6) in our sample provided evidence of multiple student learning assessment measures and could address individual and student group performance and growth.
4.1.5 Evidence of student-learning growth connected to the identified objectives.	Year 1: 3, 4, 2 Year 2: 4, 3, 4	We found no discernible patterns across grade levels and years of data collection. We observed evidence that the majority of K-12 students met expected levels of growth in the data sources from 5 out of 6 participants.
4.2.1 (Knowledge) Program completers prioritize student well-being. They maintain a “growth mindset” toward their students.	Year 1: 4, 4, 4 Year 2: 4, 4, 4	The data suggest the graduates in our sample consistently prioritize student well-being. Evidence across two data cycles and multiple data sources suggests our graduates maintain a growth mindset toward their K-12 students.
4.2.1 (Knowledge) Program completers prioritize student well-being. They maintain a	Year 1: 4, 4, 4 Year 2: 4, 4, 4	The data suggest the graduates in our sample consistently prioritize student well-being. Evidence across two data

“growth mindset” toward their students.		cycles and multiple data sources suggests our graduates maintain a growth mindset toward their K-12 students.
4.2.2 (Skills) Completers exhibit a flexible response to student needs.	Year 1: 4, 4, 4 Year 2: 4, 3, 4	All of the graduates in our sample adjusted sequence, pacing, or content based on student needs. 5 out of 6 graduates in our sample made adjustments based on the needs of individuals as well as the needs of the whole group.
4.2.3 (Dispositions) Completers demonstrate a sense of agency as an individual and as a member of learning community. They take action based on the needs of students.	Year 1: 3, 4, 4 Year 2: 4, 2, 4	5 out of the 6 graduates in our sample are members of professional communities and are actively seeking professional development. 3 of the 6 could provide examples of advocacy work based on the needs of students. The graduate who received a score of 2 was in a context where she perceived was not a supportive learning community. For example, all teachers within a specific grade level and content area were expected to do the same thing at the same time.

The scoring process provided us with more than simple scores to report to CAEP. We learned through the process what is most valuable to us as we describe our “primary traits,” and we are developing a common language across our college and with our candidates and graduates in talking about our work.

Implications—What’s next?

We choose to distinguish between assessment and evaluation. We define assessment as gathering information without making judgments while evaluation uses portions of that information to make value judgments based on specific criteria. Assessment encourages us to view broadly, to look for connectedness within a situation, to be open to whatever direction an interview or an observation may go. Evaluation focuses our attention on specific evidence and outcomes.

It was evident from the beginning that our methodological design of gathering data (assessing) and then later putting a value judgment (evaluating) on specific aspects of our data served two important groups in different ways. The professors who were involved in gathering data (assessing) saw themselves as the beneficiaries of the process, an internal stakeholder in all these efforts. They found themselves energized by the experience as they asked questions of each other and looked for connections between the data and their experiences with students in the preservice portion of our program. They informally brainstormed ways to share the positive experiences with the rest of the faculty in some way in addition to a faculty meeting.

The professors who used the data from the pre-observation interview, the classroom observation, and the post-observation interview were very focused on giving an accurate value judgment (rubric rating) based on the criteria outlined in the rubric. They were very concerned about the ultimate judgment that would come from our own external evaluator, CAEP. Especially for the first question focusing on student achievement, the evaluating professors wanted specific data. “Show me the money,” and “Cut to the chase,” were common refrains. There was an impatience with the amount

of data available. Instead of discussing points to celebrate in our graduates' work and possible adaptations for our program, they found themselves talking about ways to streamline the data-gathering process and resulting data. They were frustrated our graduates didn't simply hand us neat and tidy pre/post test scores.

Our approach to data collection and subsequent scoring of the data on a four-point scale for all the subcomponents of CAEP 4.1 and 4.2 highlighted our distinction between assessment and evaluation. We struggle to find a balance between our desire to understand in depth the effect of EPP on our graduates and their K-12 students and finding a numerical way to evaluate our graduates' impact.

Going forward, our College of Education professors and instructors will each year examine the data for general trends across all the subcomponents of CAEP 4.1 and 4.2 based on the scores. In order to have access to the richness of the data, the faculty will look specifically at the data in one case study for 4.1.4, "The teacher provides evidence of student learning," and 4.1.5, "The observer sees evidence of student-learning growth connected to the identified objectives." The resulting conversations about the scores for each subcomponent and the rich data for evidence of student learning will focus on the question "If this, then what next?" The faculty will determine next steps for program improvement.

References

Lloyd-Jones, R. (1977). Primary trait scoring. In C. R. Cooper and L. Odell (Eds.), *Evaluating writing: Describing, measuring, and judging*. New York: National Council of Teachers of English, 33-66.