Utilizing Collaborative Analysis of Student Learning in Educator Preparation Programs for Continuous Improvement

Susan Colby¹, Monica Lambert¹, and Jennifer McGee¹

Abstract
In this results-oriented era of accountability, educator preparation programs are called upon to provide comprehensive data related to student and program outcomes while also providing evidence of continuous improvement. Collaborative Analysis of Student Learning (CASL) is one approach for fostering critical inquiry about student learning. Graduate educator preparation programs in our university used collaborative analysis as the basis for continuous improvement during an accreditation cycle. As authors of this study, we sought to better understand how graduate program directors and faculty used collaborative analysis to inform practice and improve programs. Our findings suggested that CASL has the potential to foster collective responsibility for student learning, but only with a strong commitment from administrators and faculty, purposefully designed protocols and processes, fidelity to the CASL method, and a focus on professional development. Through CASL, programs have the ability to produce meaningful data related to student and program outcomes and meet the requirements for accreditation.

Keywords
educator accreditation, teacher education, program evaluation

Introduction
In this results-oriented era of accountability, educators at all levels are increasingly evaluated by their impact on student achievement. Reeves (2007) noted that educators exist in a world where too often assessment is equated with high-stakes testing. This narrow view of assessment has reached its way into the arena of educator preparation (Norman & Sherwood, 2015). Outside agencies evaluate quality of data with, at times, little correlation to student and program effectiveness (Fuller, 2014). Value-added measures that assess the performance of educator preparation programs using K-12 student data are increasingly popular (Darling-Hammond, 2015; Lincove, Osborne, Dillon, & Mills, 2014) with the potential “for good and for ill” (Harris & Herrington, 2015, p. 71). The Council for the Accreditation of Educator Preparation (CAEP) discussed these practices in their recently drafted CAEP Accreditation Handbook:

CAEP (2016) had, however, attempted to “increase the value of accreditation” (p. 5) by expanding accreditation evidence to include data related to continuous program improvement. Educator preparation providers now submit evidence of completer proficiency and evidence that a “culture of evidence” has been created that promotes “the practice of using evidence to increase the effectiveness of preparation programs” (CAEP, 2016, p. 6). Given this new reality, it is increasingly important for educator preparation providers to develop methods, tools, and processes that accomplish multiple purposes.

The journey to create a meaningful and coherent assessment system for our educator preparation programs began 3 years ago prior to our college’s most recent National Council for Accreditation of Teacher Education (NCATE) review. College administrators and program directors were challenged to develop an assessment system that would meet unit-level accreditation requirements while meeting the needs of numerous programs with discipline-specific standards and practices. Fullan and Quinn (2016) offered...
guidance for attending to these multiple demands: “Internal accountability must precede external accountability if lasting improvement in student achievement is the goal” (p. 111). Fullan and Quinn further explained this shift from compliance to collective responsibility:

Constantly improving and refining instructional practice so that students can engage in deep learning tasks is perhaps the single most important responsibility of the teaching profession and of educational systems as a whole. In this sense, accountability as defined here is not limited to mere gains in test scores but on deeper and more meaningful learning for all students.” (p. 110)

As administrators and program directors worked to create a sound assessment system, Hattie’s (2012) Visible Learning (VL) research provided a starting point.

**Searching for an Approach: The VL Framework**

Hattie’s VL research was based on a synthesis of 1,200 meta-analyses from over 50,000 individual studies. Each research study examined the influence of a program, policy, or innovation on academic achievement (Hattie, 2012, 2015). Hattie’s VL research was the largest meta-analysis ever conducted in the field of education and provided insight into what educators should, and should not, focus on in their efforts to improve practice. The key finding from this research was learning should be the explicit and primary goal and teachers need to “know thy impact” (Hattie, 2012). Educators should see their fundamental role as an evaluator of their own effect on students (Hattie, 2015). Student learning is increased (a) when teachers believe their major role is to evaluate their impact and (b) when teachers work together to know and evaluate their impact (Hattie, 2015, p. 81).

Hattie’s VL research provides guidance for educator preparation programs that not only wish to study their impact, but are also required to do so for accreditation. Hattie translated his findings for higher education: “Faculty [in higher education] need to go beyond merely collecting data, creating reports, and asking students to fill in surveys, but to become excellent interpreters of evidence about their impact” (Hattie, 2015, p. 89). Hattie’s findings had implications for our educator preparation programs. To create a culture of evidence, our college needed to build capacity in our programs so faculty would have the skills, data, and structures for studying impact (Jimerson, 2014; Katz & Dack, 2014; Schneider & Gowan, 2013; Wayman & Jimerson, 2014). One well-respected method of building capacity for professional learning is through the use of professional learning communities (PLCs).

**Establishing PLCs.** PLCs build the collective capacity of stakeholders to improve practice and have become the norm in schools since the early work of DuFour and Eaker (1998). DuFour and Eaker believed that PLCs are “the most promising strategy for sustained, substantive school improvement” (p. xi). Researchers and scholars found PLCs a promising technique to improve practice and increase student learning (Mitchell, 2014; Ronfeldt, Farmer, McQueen, & Grissom, 2015). Schmoker (2005), in the foreword of an edited book on PLCs, stated,

> If there is anything that the research community agrees on, it is this: The right kind of continuous, structured teacher collaboration improves the quality of teaching and pays big, often immediate, dividends in student learning and professional moral in virtually any setting. (p. xii)

Utilizing PLCs as a strategy to foster reflection on practice in a continuous improvement cycle seemed evident given the core principles of PLCs: (a) ensuring that students learn, (b) creating a culture of collaboration for improvement, and (c) focusing on results (DuFour, 2015; DuFour & Eaker, 1998). These core principles aligned with CAEP standards and the PLC literature provided direction for how educator preparation programs in our college could engage in collective inquiry. Ronfeldt et al. (2015) examined the impact of collaboration on student learning in a large-scale study and found while naturally occurring collaboration across a wide range of contexts is beneficial, the strongest student achievement gains resulted from collaboration focused on assessment. Van Lare and Brazer’s (2013) sociocultural model conceptualized how teachers learn in a PLC and proposed that learning must be the focus of a PLC in order to move beyond the routine outcomes of consistency and predictability. Gallimore, Ermeling, Saunders, and Goldenberg (2009) found a positive impact on student learning when settings were stable, teachers were focused on concrete learning goals, and progress was tracked. Most importantly, they found momentum increased after teachers saw the positive results of their collaboration on student learning. In the higher education arena, Mitchell (2014) found that PLCs afforded academics the ability to engage in meta-cognitive processes that helped faculty better understand the impact of their teaching on their students. These studies confirmed the importance of creating inquiry-focused protocols and structured organizational routines to guide efforts for critical, recursive reflection (Gallimore et al., 2009; Thompson, Hagenah, Lohwasser, & Laxton, 2015).

The PLC model provided our college with a sustainable structure for collaborative action; however, impact on student learning cannot be determined without sufficient skill in analyzing multiple sources of data and drawing accurate conclusions. One challenge that our college encountered was developing capacity for data-based decisions.

**Developing capacity for effective use of data.** Educational organizations struggle with similar problems when using data to inform decision making: a lack of access or ability to select appropriate data, a lack of skill in the use of data, and a lack of collaboration around the use of data (Schildkamp, Karbautzki,
Selecting a Method: The CASL

The CASL “is a professional learning design in which teacher groups analyze student work samples and assessments to learn how to effectively support students’ learning of complex academic standards” (Colton et al., 2015, p. 44). CASL framework provides guidance for instructor teams that examine the strengths and weaknesses of student responses, determine whether students meet standards and objectives, and identify implications for improvement (Langer, Colton, & Goff, 2003). Instructors look for “clues in the work” (Langer et al., 2003, p. 34) that can provide helpful information for data-informed decisions about student learning and instruction. Student work is defined as “any data or evidence teachers’ collect that reveals information about student learning (e.g., standardized test data, classroom assessments, writing samples, projects, oral reports, videotapes, pictures, or student observation data)” (Langer et al., 2003, p. 4). Critical to the reflective inquiry process is the need for a “trusting, collaborative environment” (Langer et al., 2003, p. 27) “without fear of being judged or criticized” (Colton et al., 2015, p. 40). In this supportive environment, teachers are asked to consider other viewpoints as they construct meaning that goes beyond more immediate assumptions and generalizations.

CASL is grounded in decades of work fostering reflective and collaborative inquiry in teachers as a means to improve student learning. CASL research began in 1986 with graduate students (Langer et al., 2003; Sparks-Langer, Simmons, Pasch, Colton, & Starko, 1990) and progressed to work in public schools. CASL had strong connections with peer coaching models (Costa & Garmston, 2002; Stallings, Needels, & Sparks, 1987). The insights learned from CASL were used to inform the National Board for Professional Teaching Standards (NBPTS) and for professional development for public school teachers. At the classroom level, CASL can assist teachers in interpreting student learning evidenced in assessments, identifying areas for improvement, and reflecting on teaching practice. At the program/school level, CASL can be used to determine collective student learning; examine alignment among curriculum, instruction, and assessment; and identify areas for program improvement (Sparks-Langer et al., 1990). This reflective thinking results in a clearer vision of needed improvements (Sparks-Langer et al., 1990).

Purpose of the Study

The purpose of this study was to examine how collaborative analysis was used in our college’s graduate accredited programs for continuous program improvement. The researchers sought to understand how graduate programs used results from collaborative analysis to inform practice and what facilitators of collaborative analysis sessions learned from the process. The research questions for this study were the following:

Research Question 1: How do program directors and program faculty use results from CASL to inform practice in a continuous improvement cycle?

Research Question 2: What can we learn from the experiences of program directors who have facilitated collaborative analysis sessions?

This research study and the corresponding protocols received approval from our Office of Research Protections Internal Review Board.

Method

This qualitative research study was conducted in a college of education at a midsize comprehensive university with a large educator preparation program that graduates approximately
500 undergraduates and 285 graduates per year. Programs at both the undergraduate and graduate level are NCATE approved. Our college is currently transitioning to CAEP standards and will undergo CAEP accreditation in our next cycle. One CAEP principle in particular provided an impetus for our work, “There must be solid evidence that the provider’s educator staff have the capacity to create a culture of evidence and use it to maintain and enhance the quality of the professional programs they offer” (CAEP, 2015, Introduction). To more closely align CASL with the multiple measures requirement of CAEP, the original student work definition was broadened to include collaborative analysis of any appropriate data source (e.g. survey data, enrollment and graduation numbers, graduate student evaluations).

Research Design

Qualitative researchers typically gather data from multiple sources to help ensure accurate results. Researchers “review all of the data, make sense of it, and organize it into categories or themes that cut across all of the data sources” (Creswell, 2014, pp. 185-186). The research team for this study, also the authors of this article, consisted of two college administrators and one former director of teacher education assessment. The team analyzed documents and survey responses from graduate program directors to establish emerging themes and draw conclusions related to the research questions. For the document review, the team analyzed assessment reflections from graduate program directors about collaborative analysis sessions. Assessment reflections provided documentation of continuous program improvement with program area faculty for accreditation purposes. Assessment reflections served as the primary data source for this study. Creswell (2014) noted that there are advantages and limitations to using each type of data in qualitative research. An advantage of document review is that the original language becomes the source for analysis. Documents have limitations, however, and participants may not be equally articulate or perceptive. Further, a document may not comprehensively capture an event (Creswell, 2014). Survey responses to open-ended questions from program directors about their experiences facilitating collaborative analysis sessions served as a secondary data source for this study. The research team found the advantages of survey research to be particularly helpful in our study due to the economy of the design and the rapid turnaround in data collection (Creswell, 2014). For this study, researchers were able to sample program directors efficiently about their experiences at the close of the assessment reflection cycles.

Participants

The participants in this study (n = 13) were graduate program directors from NCATE accredited educator preparation programs. A single-stage sampling procedure was utilized as the researchers had access to all of the names in the population and were able to access the documents and participant perceptions directly (Creswell, 2014). These graduate programs included Curriculum Specialist, Elementary Education, Higher Education, Educational Media and Technology, Library Science, Mathematics Education, Middle Grades Education, Music Education, Professional School Counseling, Reading Education, Romance Languages, School Administration, and Special Education. Program directors were tenure-track faculty selected by departmental chairs and affiliated program faculty due to their program expertise and administrative skills. Program directors attended college level meetings and worked closely with departmental chairs and deans to manage programs. Program area faculty, while not direct participants in this study, did participate in collaborative analysis sessions which served as the basis of the assessment reflections. Program area faculty were primarily tenure-track faculty that taught courses in the program and participated in program area meetings. Program directors and program faculty met regularly to discuss advising, curriculum, program assessments, and other program-related topics. Participants had no prior knowledge of CASL techniques and most programs did not routinely review student work for program improvement purposes.

Instruments

The data collected and analyzed for this study included assessment reflections based on collaborative analysis sessions and survey responses from program directors describing their experiences as facilitators of collaborative sessions. The assessment reflection was developed by the director of teacher education assessment. The assessment reflection template included a series of prompts to assist programs in attending to the various stages in the collaborative analysis process. The survey for program directors was developed by the authors of this study. The survey was cross-sectional and data were collected electronically using university approved survey software for research. Both the assessment reflection template and the survey for program directors were approved by Institutional Review Board (IRB).

Assessment reflections. Each program director was asked to complete two assessment reflections based on collaborative analysis sessions with faculty during the 2014-2015 academic year. The assessment reflection template was designed to stimulate collaborative inquiry using multiple sources of data and to provide a framework for the written reflection. Guiding questions assisted the program director in leading a discussion around five key areas. For the first section, program directors were asked to list the action item(s) for the session. Program directors identified topics for discussion that were based on program priorities. In the second section, program directors reported on the status of past improvements. Program directors were asked to describe the status of changes that were made to the program as a result of recent reviews. In
the third section, program directors cited the data sources that would be reviewed in the current session. These data sources were to align with the action items. For the fourth section, program directors described how data were interpreted. Directors were to report on what was learned about student performance and how the curriculum impacts student performance based on a review of data. In addition, program directors were to identify which program goals and objectives findings applied. In the final section of the written reflection, program directors responded to a series of questions about program improvement. Directors were asked to describe the changes that the program would make based on the analysis of data. Directors were prompted to specifically describe changes to the curriculum and the assessment system. Directors were also asked to provide detail on the next steps.

The assessment reflections were discussed in program director meetings in the fall of 2014. The dean of the college expressed the need for programs to use a common data source for providing evidence of continuous improvement and introduced the assessment reflection template. Program directors were informed that data would be aggregated across programs. Program faculty were to engage in multiple cycles of collaborative analysis for continuous improvement purposes.

**Survey responses.** Survey responses from program directors about their experiences as facilitators of collaborative analysis sessions served as the secondary data source for this study. Survey responses were anonymous. There were five open-ended survey items related to facilitating collaborative analysis. The first question asked program directors if they found collaborative analysis a useful process. They were asked to provide a rationale for their response. The second question asked program directors if the prompts provided in the assessment reflection proved useful in guiding the collaborative discussions. Again, directors were asked to provide a rationale for their answer. The third question prompted directors to describe the level of difficulty with conducting collaborative analysis sessions. Program directors, for the fourth question, were to describe the challenges they faced in leading a group of faculty through this process. Last, directors were asked for their suggestions on what would help them facilitate the process better in the future.

Surveys were sent electronically to program directors in early spring 2016 after two cycles of collaborative analysis sessions. The introduction to the survey included a description of the research study, the purpose for the study, the researchers involved, and an overview of the survey. The survey overview and questions were IRB approved. Program directors were given a choice to opt out of the survey. The survey remained open for 2 weeks.

**Data Analysis**

Creswell’s (2014) six-step process for analyzing qualitative data was used as a guiding framework: (a) organizing and preparing the data for analysis, (b) reading for a general sense of the information, (c) coding of all data, (d) identifying themes from the coding and searching for theme connections, (e) representing the data, and (f) interpreting the larger meaning of the data. The researcher’s purpose was to determine how faculty were using collaborative analysis to reflect on student learning and determine areas for improvement. In addition, the researchers hoped to better understand the experiences of program directors as facilitators of collaborative analysis sessions.

During Phase 1 of data analysis, the researchers analyzed the assessment reflections. First, each researcher independently open-coded the assessment reflections from the fall of 2014. The research team met on several occasions during the coding process to discuss findings and to cross-check developing codes. Once individual coding was completed, the research team met to discuss findings and create a list of codes that appeared across the fall assessment reflection narratives. Assessment reflections from the spring of 2015 were analyzed using the same process. The research team met to refine the codes and check for themes emerging from both sets of data.

In Phase 2, the same process was followed for analyzing the survey data. The researchers each independently coded the survey data collected from anonymous program directors. The research team then met to determine the emerging patterns evident in the set of responses.

After all three data sources were coded, the researchers met to discuss emergent themes across the data collection points. The researchers finalized the primary themes related to the research questions and developed conclusions and implications for the study.

It is well known that unconscious bias can be a limitation or source of error when analyzing and interpreting data. Few researchers achieve complete objectivity (Best & Kahn, 1998). Because this study was primarily qualitative, it was important for the researchers to examine their own bias and potential influence on the study. The research team incorporated multiple strategies to enhance the accuracy of the findings. These strategies included triangulation, consideration of bias, reporting of negative and discrepant information, and cross-checking of codes during analysis (Creswell, 2014). The research team collected data from two data sources to develop emerging themes and build a cohesive interpretation of the data. Because the assessment reflection template was created by the current assessment director with no input from the research team, bias was not a factor in the design of the template or in the responses recorded in the assessment reflections. In addition, program directors were required to complete assessment reflections, and participation rate was not influenced by the researchers. To mitigate researcher bias during the analysis, assessment reflections were not reviewed until both cycles were completed. Furthermore, the research team met on several occasions during the coding process to ensure interrater reliability by
cross-checking codes and reconciling discrepancies. The research team did, however, develop the survey and therefore greater potential existed for influencing the data and the interpretation of the data. The research team disclosed the purpose of the survey in the survey introduction and highlighted that there were no consequences for choosing not to participate.

Results

In total, 21 collaborative analysis sessions in 13 graduate programs were held over the course of the academic year. Program directors submitted assessment reflections for fall 2014 (n = 13) and spring 2015 (n = 8). The completion rate for the assessment reflections was higher in the fall as a result of a stronger effort for compliance. The dean sent out multiple reminders in the fall to program directors to complete the assessment reflections. The number of program faculty participating in the collaborative analysis sessions ranged from two to 18 depending on the size of the program. After coding the assessment reflections and identifying emerging themes, the research team found it helpful to note frequencies for specific topics discussed, types of data used in analysis, and action steps identified by program faculty.

Surveys were sent electronically to the 13 graduate program directors after all collaborative analysis sessions were conducted. Two reminders were sent by the associate dean, one of the researchers for this study. Six program directors completed the anonymous survey. Seven program directors opted out of the survey. Surveys were coded after a 2-week open window. The researchers were aware of the potential for nonresponders to influence the overall survey data (Creswell, 2014).

Findings were categorized into three primary themes: Outcomes of Collaborative Analysis, Use of Data for Collaborative Analysis, and Facilitating Collaborative Analysis. These themes aligned with the research questions. Findings from the analysis of the assessment reflections primarily informed Research Question 1: How do program directors and program faculty use results from CASL to inform practice in a continuous improvement cycle? Findings from the analysis of the survey responses primarily informed Research Question 2: What can we learn from the experiences of program directors who have facilitated collaborative analysis sessions?

Outcomes of Collaborative Analysis

Each assessment reflection opened with the action item(s) for the collaborative analysis session. Action items were specific and served as the focus for the session. A survey of the action item topics revealed that programs had conversations around course development, curricular changes, and student products. Action items separated into two categories: items related to program goals and items related to student learning outcomes (SLOs). Program-related action items included aligning documents and protocols to standards, revising assignment guidelines and rubrics, and examining and approving unit-level standardized rubrics. Student learning action items included implementing strategies for promoting deeper critical thinking, integrating skills in student products, and addressing perceived student deficiencies.

Programs conducted collaborative analysis sessions using a variety of data sources aligned with the action item. Use of data is explained in the second theme below. Program directors and faculty reached consensus at the conclusion of each session about recommendations for improvement. These recommendations were recorded in the reflections. A review of the continuous improvement section of the assessment reflections revealed that programs identified a wide range of action steps. A majority of the programs intended to review and revise current rubrics or create new rubrics. These programs cited a need for continuous monitoring of goals. In addition, many programs realized consensus was needed by program faculty on course expectations, assignments, and/or evaluations. Specific actions included revisions to assignments, rubrics, and courses. A few programs realized they lacked important program documents such as curriculum maps or assessment plans and action steps were included for the development of these program documents.

Use of Data for Collaborative Analysis

The prompts on the assessment reflection template were intended to elicit group discussions about student learning and program quality. Program directors were asked to cite the data sources used for the review and describe what was learned through collaborative analysis. Programs varied greatly in the type of data used for collaborative analysis and how data were interpreted.

Data sources cited. A review of the 21 assessment reflections revealed that a total of 37 data sources were identified throughout the reports. These data represented five types: student assessments (62%), graduate student exit surveys and interviews (19%), evaluations of faculty and supervisors (8%), documents (5%), and other (5%). Data from student assessments were predominantly used to make decisions about program improvement.

From the student assessment category, the assessment most often cited for analysis was the Product of Learning (POL). The POL was the capstone assessment for all NCATE accredited graduate programs and consisted of a collection of artifacts and accompanying synthesis reflections providing evidence that five state standards were met. The POL rubric was common across all programs, and results from each criterion on the POL rubric were aggregated for unit-level accreditation. Although the rubric for all programs was the same, the contents of the portfolio and the synthesis reflections varied depending on program. Each program had the
latitude to create course and program assignments that met specific objectives. A second source of assessment data often used in the collaborative analysis sessions was the depth of content evidence. This evidence was also required at the state level and each graduate program required students to demonstrate discipline-specific depth of content. A third student assessment data source cited was capstone course assessments. Multiple program directors used data collected from rubrics to determine next steps for improvement. Other assessments, such as comprehensive exams, portfolios, and term papers, were cited less frequently.

It was evident that programs relied heavily on aggregated quantitative data when discussing student learning. These data were gathered from ratings on proficiency levels (does not meet, meets, exceeds expectations) for specific criteria from the assessment rubrics. Nearly all programs analyzed data from multiple criteria on the rubric. Most programs used this quantitative data without examining individual student work to better understand the results. For example, one graduate program (GP 4) stated, “Our data from the previous semester demonstrates that most graduate students are performing at ‘proficient’ or better on all assessment points of their Products of Learning (2.49-2.74/4).” A second graduate program (GP 2) stated, “We achieved the first of our two criteria for graduate student demonstration of depth of content knowledge with 100 percent of our students (n = 19) achieving a score of 8 or higher, indicating proficient performance.” This type of reporting shows a cursory glance at aggregated data devoid of an analysis of student work. This was true for nearly all programs that used assessment data. Only one graduate program (GP 11) indicated that they had reviewed individual student products:

In previous Product of Learning [final project] presentations we have noticed that non-native speakers of the target language do not always have a strategy in place to make ongoing, independent progress in their target language skills outside of our graduate classes, thus they struggle somewhat in the completion of target language papers, oral presentations, or both.

This graduate program reviewed student work from multiple assignments to determine what students were learning. Program faculty found a difference in the use of target language between graduate students with and without international experiences. The program now strongly recommends a variety of international experiences for second-language learners to enhance their target language competence.

A second category of data sources was graduate student perceptions collected from surveys and interviews. These data provided programs with rich descriptions about program strengths, weaknesses, and areas for improvement. Programs relied on survey and interview data to provide confirmation and triangulation. For example, one graduate program (GP 2) used survey responses to assist in developing a program of study for an accelerated admissions program. Data provided the program with feedback about which courses were not appropriate for the condensed summer schedule and which courses were not appropriate for a double-count (undergraduate and graduate). Student perceptions also revealed that more technology integration was needed throughout the course work. The exit survey data from a second graduate program (GP 9) revealed that graduate students needed earlier exposure to specific skills in order for mastery by graduation.

A third category of data related to faculty. These data included quantitative and qualitative data from course evaluations of teaching, field supervisor evaluations, and faculty perceptions related to courses and/or program. One graduate program (GP 6) used end of course evaluations from a new course to offer student perceptions about course effectiveness. The qualitative comments were helpful in determining what students were gaining from the course and how the course could be refined to better meet student needs. A second graduate program (GP 9) used graduate student evaluations of university and school-based field supervisors to evaluate the program. In this program, high-quality field experiences are crucial to gaining knowledge and skills required in the field. Data were used to determine whether supervisors were effective in their positions and if changes needed to be made in the selection and mentoring of supervisors. In addition, one graduate program (GP 7) used faculty (instructor) perceptions about the quality of writing from specific course assignments to make recommendations about how to improve the quality of student writing throughout the program.

The fourth and fifth data categories, much less frequently cited, referred to a review of documents and grades. A total of four programs used these two data sources to determine areas for improvement. Interestingly, neither of the graduate programs that reviewed grades (GP 1 and GP 12) found student grades useful in their deliberations about student learning. Faculty in two graduate programs (GP 3 and GP 5) analyzed the alignment of program documents to state and/or national standards. Both of these programs did report areas for improvement based on their review of documents.

**Interpretation of data.** All program directors were able to describe how they interpreted meaning from the data sources their programs reviewed. There was strong alignment between the sources cited and the interpretation of these data. A review of the 21 assessment reflections revealed a total of 51 specific findings about how data were interpreted and would be used for program improvement. These findings separated into four categories and were related to the following: assessments (62%), program improvement (21%), supporting students (12%), and faculty (5%).

Nearly all comments in the interpretation section described findings related to assessments. This was not surprising given that a majority of data reviewed were from program assessments. Half of the interpretive comments
reported proficiency levels of students on course assessments and satisfaction with the percentage of graduate students rated as meeting and exceeding specific criteria. The second half of assessment related comments focused on rubrics. Program directors explained how data provided evidence that rubrics needed refining. Typically, rubric revisions were needed to yield more variance or more specificity in the descriptors. A few directors noted that there was duplication in rubrics, and one director reported that a large amount of missing data made it difficult to determine next steps.

A second set of findings were related to the overall program. Nine program directors recommended changes in the program of study or in a specific course. A few directors noticed there was duplication in courses while others noted courses that were no longer useful based on changing standards and practices in the field. A couple of directors stated that new data sources were needed to better make decisions about program improvement. One director noted that there was a large discrepancy between graduate student self-report and faculty ratings on positions. This program intended to learn more about this disparity.

A third set of findings focused on students. Programs understood the need to develop and implement formative checks for capstone portfolios and products. A couple of program directors identified a need for more scaffolding, support, and remediation of students enrolled in programs. One program found a need to better meet the differences of different student groups.

The last set of findings were related to faculty. After reviewing various data sources, programs learned that faculty needed to be more consistent in their teaching and when grading program assessments.

**Conclusions about use of data.** After reviewing all of the assessment reflections for findings related to the use of data, it was evident that programs consulted a wide variety of data to make decisions about program improvement. As evidenced in the first assessment reflections, the focus of many collaborative analysis sessions shifted from a discussion about student learning to a discussion about the program. Implications from a review of these quantitative rubric ratings generally resulted in recommendations about revising the rubric with little confirmation from the examination of student work. Only one program analyzed actual student work. In general, programs selected important program assessments and appropriate data sources, but nearly always made their decisions about how to improve their program based on quantitative scores rather than rich, descriptive qualitative findings that emerged from examining student work.

Programs that participated in two collaborative analysis sessions appeared to have stronger results in the second session. They posed fewer questions and used less data sources but produced deeper reflections. One program director reflected on this progression in the survey: “It [the collaborative analysis session] was difficult both times, but easier the second time, mainly because I had a better idea of how to use it more to my and the program’s advantage.”

**Facilitating Collaborative Analysis**

Program directors were asked to complete a survey about their experiences as a facilitator. The first question asked directors whether collaborative analysis was a useful process. Overwhelmingly, directors indicated that collaborative analysis sessions were useful or somewhat useful. Program director (PD 4) discussed the benefits:

The process forced us to take a positivist look at the progress/success of our students according to our own metrics and analyze just what those metrics were telling us. We had spirited debate about (1) the content of our courses and entire curriculum, (2) the sequencing of our courses, (3) our teaching styles and methods, (4) the quality of our students, (5) the support mechanisms in place—and whether they were utilized or even appropriate—to help develop the areas that we believed were necessary for success, and (6) whether we need to make adjustments to any of these aspects, including the curriculum. We even took the time to do a bit of an environmental scan to assess whether there were changes in the profession requiring an adjustment in our curriculum, including needing new courses.

Only one program director (PD 3) reported “not particularly” useful when responding to this question. This director stated, “Because the faculty meet to discuss program goals and curriculum twice a semester already, this really did not add to what we already do—so I guess you could say it was useful (but not a new thing).”

The second survey question asked directors about the helpfulness of the assessment reflection prompts. A majority of program directors found the prompts helpful. One program director (PD 1) stated, “I felt that I was starting from scratch with a new program and the prompts were very helpful.” A second program director (PD 5) stated, “They got a bit tedious and detailed, but ultimately I have to admit they were valuable as they drilled down.” It does appear that the prompts assisted in providing a focus for the conversations, especially when directors followed the sequence outlined in the template.

The third question in the survey asked program directors about the level of difficulty in conducting collaborative analysis with program faculty. The responses were mixed with half of the program directors stating the process was easy. Alternatively, two program directors (PD 5 and 6) found the process difficult because they had to select and organize the data for examination. Faculty played no role in preparing for the session. These mixed responses were not surprising given the various comfort levels in leading data-driven discussions.

The fourth survey question asked program directors to describe the challenges they faced while leading faculty through the collaborative analysis process. Program
directors provided a variety of responses to this question. Two program directors (PD 5 and 6) found challenges in coming to agreement on ratings, metrics, and principles. A second program director (PD 2) stated, “We struggle with determining how to be sure we are uniformly applying the ratings in the POL rubric so that the variability in the results is really related to the program outcomes and not the raters.” Two program directors (PD 4 and 6) reported a lack of interest from faculty and PD 4 reported faculty hostility toward the process.

The final question asked program directors what would help facilitate the collaborative analysis process. Five directors responded to this question with over half discussing the need for more faculty engagement. This was evident in the following response (PD 1):

I think it would be extremely beneficial if EVERYONE had more input and responsibility in this process, not just the program directors. Had I had input prior to becoming program director, I might have been less lost and may have seen the value in the process. It would seem assessment should be part of everyone’s responsibility to providing/maintaining quality programs for our students and excellent teachers for our state and beyond.

A second program director (PD 4) confirmed the need for faculty input:

Much of this necessarily falls on Program Directors. We are the only ones with access to the data at the program level; it falls to us to pull that data and make it accessible to faculty. It turns more into a presentation of possible focus areas and assent by other faculty.

A third program director added insight about the role of administrators in the assessment process (PD 6):

Program faculty need to see program improvement and assessment as a PROPERTY of teaching at a university. As long as that message is disseminated only to directors, directors will bear the brunt of assessment for the entire program. Program faculty need to hear this message from administration—at all levels.

The results indicated that more ownership around assessment needs to be cultivated in our college, and administrators need to find ways to support program directors as they lead faculty through this process.

**Implications**

From the three themes that emerged from this study, Outcomes of Collaborative Analysis, Use of Data for Collaborative Analysis, and Facilitating Collaborative Analysis, four implications for implementing collaborative analysis as a component of a unit assessment system were identified. These implications are discussed below.

**Commitment is Crucial**

One implication that emerged from this study was the need for a strong commitment from administrators, program directors, and faculty for successful implementation of the collaborative analysis process. Efforts to implement collaborative analysis must be systematic and continuous, especially if data are to be aggregated for unit-level results. Too often, data analyses are not conducted frequently enough or in enough depth to reap the beneficial results of formative assessment (Hoover & Abrams, 2013). To use collaborative analysis for accreditation, all programs should engage in collaboration on an ongoing basis to maximize the potential for significant change.

Collaborative analysis should become a primary focus for improvement purposes rather than an ancillary activity. Fullan and Quinn (2016) stated that it is easy to create fragmentation when responding to multiple mandates for accountability and improvement. They propose that to build coherence in a system, the right drivers must be set in action. Too often, the wrong drivers of punitive accountability, individualistic strategies, technology, and ad hoc policies drive efforts for reform. Utilizing the right drivers of focused direction, cultivating collaborative cultures, securing accountability, and deepening learning will help organizations “become immersed in work that develops focus and coherence across systems” (Fullan & Quinn, 2016, p. 11).

Findings from this study suggested that program directors struggled, at times, to implement collaborative analysis because faculty investment and administrative direction were lacking. As researchers for this study, we learned that a stronger commitment to the process and a strategic plan for fostering an investment from program faculty would likely have created more coherency and commitment.

**Implement CASL With Fidelity**

A second implication that emerged from this study was the importance of implementing CASL with fidelity to the original model. CASL is not the only approach for focusing teams on collaborative analysis of data, but it is a comprehensive model that has been refined for over a decade (Colton et al., 2015; Langer et al., 2003). Findings suggested that our programs would have benefited from stricter adherence to the original definition of student work and to the reflective inquiry phases. Data from this study provided evidence that programs tended to revert back to familiar patterns for examining data in lieu of a more formal process for analyzing student work. This practice is noted in the literature (Gallimore et al., 2009; Hoover & Abrams, 2013; Thompson et al., 2015; Van Lare & Brazer, 2013). Nearly half of the graduate programs examined aggregated quantitative results from capstone projects to determine what percentage of graduate students were proficient. These data were used to identify areas for program improvement without confirmation from
student work. In addition, many of our programs used data that were only indirectly related to student learning. Programs used faculty and student perceptions, disposition data, documents, and grades to determine areas for improvement. While we acknowledge that examining multiple sources of data is beneficial, the process established in our college did not provide the impetus for examining student work that would help faculty “reexamine, clarify, and transform their thinking so that they can help students succeed” (Langer et al., 2003, p. 13). Although the template was designed to foster collaborative discussion, it did not stimulate the type of reflective inquiry that provides a deep understanding of student learning as outlined in the CASL method. As Fullan and Quinn (2016) suggested, “we must shift from a focus on teaching or inputs to a deeper understanding of the process of learning and how we can influence it” (p. 79). This does not mean that programs should only review student work in collaborative sessions; indeed, collaboration of all types has potential for improving student learning (Ronfeldt et al., 2015). However, in a continuous improvement cycle, there is a place for critical discussions based on examining real student work. True CASL sessions should have a place in an assessment system where program faculty analyze student work resulting from course and program assessments to critically reflect on practice.

**Design Protocols and Processes for Critical Inquiry**

A third implication for educators intending to implement CASL is to thoughtfully design inquiry-based protocols and processes. As evidenced in the literature, it is critical that faculty and administrators create protocols that provide the structure for focused opportunities that support reflection on practice. The assessment reflection template used in our college lacked questions that would support the effective use of data (Schildkamp et al., 2014) and stimulate critical inquiry (Gallimore et al., 2009; Langer et al., 2003). Our preparation programs might have benefited from a more focused attempt to develop social routines in our faculty that fostered productive conversations about teaching and learning that were “intellectually ambitious” (Thompson et al., 2015, p. 364). The prompts helped program directors facilitate a data analysis meeting, but they did not cultivate critical reflection about student learning. This was evidenced by some of the more perfunctory reports we received. This was also evident in the survey results. Some program directors found the prompts to be useful while others found them to be a restrictive. In addition, the assessment reflection template could have been deliberately designed to more directly align with continuous improvement as outlined in CAEP Standard 5. Including prompts that assisted in tracking results over time (5.3), examining completer impact on K–12 student growth (5.4), and including a variety of stakeholders in evaluation (5.5) would have benefited our data collection.

As a result of this study, we learned how important it is to establish SLOs prior to undertaking CASL. Better defined SLOs in programs and deliberate attention to using SLOs in protocols and processes would have produced more focused and productive CASL sessions. Fullan and Quinn (2016) confirmed the need for SLOs to become the foundation of collaborative work:

> The first step in building precision and consistent practices is to be clear about the learning goals. For the last quarter century, education has been giving superficial lip service to 21st century skills without much concerted action or impact. The energy has been invested in describing sets of skills without much robust implementation or effective ways to measure them. (p. 83)

**Address Professional Learning Needs**

A fourth implication that emerged from this study was that guidance, support, and resources must be provided to program directors and faculty implementing CASL. Addressing the needs of directors and faculty is crucial for success. Program directors need assistance in selecting appropriate data to analyze and in leading collaborative discussions. In addition, directors and faculty need to develop data-based decision-making skills. This is not surprising given the research in this area. Teachers find it difficult to interpret evidence and use data to draw accurate conclusions (Ronka, Lachat, Slaughter, & Meltzer, 2009; Schildkamp et al., 2014; Schneider & Andrade, 2013), and professional development is critical for learning how to use data for improvement purposes (Buhle & Blachowicz, 2009; Schildkamp et al., 2014; Schneider & Gowan, 2013). Modeling and coaching are highly successful professional learning strategies (Aguilar, 2013), and program directors would benefit from participating in CASL at the unit level before facilitating sessions with their faculty. Finally, a data expert can help program directors select appropriate data and provide a focus for collaborative sessions (Schildkamp et al., 2014).

**Conclusion**

The findings from this study provide insight into the potential of CASL for program improvement that supports internal and external accountability (Fullan & Quinn, 2016). The CASL process can foster collective responsibility for student learning, but only with a strong commitment from administrators and faculty, purposefully designed protocols and processes, fidelity to the program, and a focus on professional development. Faculty become engaged in critical reflection about student learning with the right structures, routines, and tools. Through CASL, programs have the ability to produce meaningful data related to student learning and program improvement. Findings from this study confirm that using any method, strategy, or system for both internal and external evaluation, while promising, has its challenges. Norman and
Sherwood (2015) confirmed this finding: “While we advocate a model of program improvement that embraces both internal and external evaluation components, such a model is not without challenge” (p. 20).

It is our hope that CASL will become a method used in our college for both program improvement that leads to greater student learning and evidence of engagement in a continuous improvement cycle for accreditation purposes. Perhaps as we prepare for our first CAEP accreditation, we will come to understand that “securing accountability is not about pleasing the system (although there is nothing wrong with this) but about acting in ways that are in your own best interest” (Fullan & Quinn, 2016). The new CAEP standards embrace some of the same principles for continuous improvement that this college values. CAEP allows for institutions to create meaningful assessment systems that support student learning, program quality, and continuous improvement. We believe that CASL, or other well-defined methods for critical conversation, can provide coherency to a college-level assessment system. As Fullan and Quinn (2016) stated, “The key to a capacity building approach lies in developing a common knowledge and skill base across all leaders and educators in the system, focusing on a few goals, and sustaining an intense effort over multiple years” (p. 57). Perhaps it is time for educator preparation colleges to more directly “know thy impact” (Hattie, 2015) and take their place in “building communities of actors whose collective work is aimed at the improvement of teaching for all students” (Thompson et al., 2015, p. 365).

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research and/or authorship of this article.

References
Aguilar, E. (2013). The art of coaching: Effective strategies for school transformation. San Francisco, CA: Jossey-Bass.
Best, J. W., & Kahn, J. V. (1998). Research in education (8th ed.).Needham Heights, MA: Allyn & Bacon.
Buhle, R., & Blachowicz, C. L. Z. (2009). The assessment plan. Educational Leadership, 66(4), 12-17.
Coburn, C. E., & Talbert, J. E. (2006). Conceptions of evidence use in school districts: Mapping the terrain. American Journal of Education, 112, 469-495.
Colton, A. B., Langer, G. M., & Goff, L. S. (2015). Create a safe space to learn. Journal of Staff Development, 36(2), 40-45.
Costa, A. L., & Garmston, R. J. (2002). Cognitive coaching: A foundation for renaissance schools (2nd ed.). Norwood, MA: Christopher-Gordon Publishers.
Council for the Accreditation of Educator Preparation. (2015). CAEP standards introduction. Available from http://caepnet.org
Council for the Accreditation of Educator Preparation. (2016). CAEP accreditation handbook. Available from http://caepnet.org
Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.). Thousand Oaks, CA: SAGE.
Darling-Hammond, L. (2015). Can value added add value to teacher evaluation? Educational Researcher, 44, 132-137.
DuFour, R. (2015). How PLCs do data right. Educational Leadership, 73(3), 22-26.
DuFour, R., & Eaker, R. E. (1998). Professional learning communities at work: Best practices for enhancing student achievement. Bloomington, IN: National Education Service.
Fullan, M., & Quinn, J. (2016). Coherence. Thousand Oaks, CA: Corwin Press.
Fuller, E. J. (2014). Shaky methods, shaky motives: A critique of the National Council of Teacher Quality’s review of teacher preparation programs. Journal of Teacher Education, 65, 63-77.
Gallimore, R., Ermeling, B. A., Saunders, W. M., & Goldenberg, C. (2009). Moving the learning of teaching closer to practice: Teacher education implications of school-based inquiry teams. The Elementary School Journal, 109, 537-553.
Harris, E. N., & Herrington, C. D. (2015). Editors’ introduction: The use of teacher value-added measures in schools: New evidence, unanswered questions, and future prospects. Educational Researcher, 44, 71-76.
Hattie, J. A. C. (2012). Visible learning for teachers. Maximizing impact on achievement. London, England: Routledge.
Hattie, J. A. C. (2015). The applicability of Visible Learning to higher education. Scholarship of Teaching and Learning in Psychology, 1, 79-91.
Hoover, N. R., & Abrams, L. M. (2013). Teachers’ instructional use of summative student assessment data. Applied Measurement in Education, 26, 219-231.
Jimerson, J. B. (2014). Thinking about data: Exploring the development of mental models for “data use” among teachers and school leaders. Studies in Educational Evaluation, 42, 5-14.
Katz, S., & Dack, L. A. (2014). Towards a culture of inquiry for data use in schools: Breaking down professional learning barriers through intentional interruption. Studies in Educational Evaluation, 42, 35-40.
Langer, G., Colton, A., & Goff, L. (2003). Collaborative analysis of student work: Improving teaching and learning. Alexandria, VA: Association for Supervision and Curriculum Development.
Lincove, J. A., Osborne, C., Dillon, A., & Mills, N. (2014). The politics and statistics of value-added modeling for accountability of teacher preparation programs. Journal of Teacher Education, 65, 24-38.
Marsh, J. A., & Farrell, C. C. (2015). How leaders can support teachers with data-driven decision making: A framework for understanding capacity building. Educational Management Administration & Leadership, 43, 269-289.
Mitchell, J. (2014). Academics learning a new language: Developing communities of practice in faculty professional development. International Journal of Adult, Community and Professional Learning, 21(1), 1-13.
Norman, P. J., & Sherwood, S. A. S. (2015). Using internal and external evaluation to shape teacher preparation curriculum: A model for continuous program improvement. The New Educator, 11(4), 4-23.
Reeves, D. (2007). From the bell curve to the mountain: A new vision for achievement, assessment, and equity. In D. Reeves (Ed.), *Ahead of the curve: The power of assessment to transform teaching and learning* (pp. 1-14). Bloomington, IN: Solution Tree.

Ronfeldt, M., Farmer, S. O., McQueen, K., & Grissom, J. A. (2015). Teacher collaboration in instructional teams and student achievement. *American Educational Research Journal, 52*, 475-514.

Ronka, D., Lachat, M. A., Slaughter, R., & Meltzer, J. (2009). Answering the questions that count. *Educational Leadership, 66*(4), 18-24.

Schildkamp, K., Karbautzki, L., & Vanhoof, J. (2014). Exploring data use practices around Europe: Identifying enablers and barriers. *Studies in Educational Evaluation, 42*, 15-24.

Schmoker, M. (2005). Here and now: Improving teaching and learning. In R. Dufour, R. Eaker, & R. Dufour (Eds.), *On common ground: The power of professional learning communities* (pp. xi-xvi). Bloomington, IN: Solution Tree Press.

Schneider, M. C., & Andrade, H. (2013). Teachers’ and administrators’ use of evidence of student learning to take action: Conclusions drawn from a special issue on formative assessment. *Applied Measurement in Education, 26*, 159-162.

Schneider, M. C., & Gowan, P. (2013). Investigating teachers’ skills in interpreting evidence of student learning. *Applied Measurement in Education, 26*, 191-204.

Sparks-Langer, G. M., Simmons, J. M., Pasch, M., Colton, A., & Starko, A. (1990). Reflective pedagogical thinking: How can we promote it and measure it? *Journal of Teacher Education, 41*(4), 23-32.

Stallings, J., Needels, M., & Sparks, G. (1987). Observation for the improvement of teaching. In D. Berliner & B. Rosenshine (Eds.), *Talks to teachers* (pp. 127-158). New York, NY: Random House.

Staman, L., Visscher, A. J., & Luyten, H. (2014). The effects of professional development on the attitudes, knowledge and skills for data-driven decision making. *Studies in Educational Evaluation, 42*, 79-90.

Thompson, J., Hagenah, S., Lohwasser, K., & Laxton, K. (2015). Problems without ceilings: How mentors and novices frame and work on problems-of-practice. *Journal of Teacher Education, 66*, 363-381.

Van Lare, M. D., & Brazer, S. D. (2013). Analyzing learning in professional learning communities: A conceptual framework. *Leadership and Policy in Schools, 12*, 374-396.

Wayman, J. C., & Jimerson, J. B. (2014). Teacher needs for data-related professional learning. *Studies in Educational Evaluation, 42*, 25-34.

Wayman, J. C., & Stringfield, S. (2006). Data use for school improvement: School practices and research perspectives. *American Journal of Education, 112*, 463-468.

Young, V. M. (2006). Teachers’ use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education, 112*, 521-548.

**Author Biographies**

**Dr. Susan Colby** is Professor and Chair in the Department of Curriculum and Instruction at Appalachian State University. Her research interests include assessment of student learning, teacher preparation program effectiveness, and the evaluation of preservice and practicing teachers. She has published in journals such as *Studies in Educational Evaluation, Action in Teacher Education, Reading Research and Instruction, Teaching Education,* and Educational Leadership.

**Dr. Monica Lambert** is Professor and Associate Dean in the Reich College of Education at Appalachian State University. Her research interests include effective P-12 partnerships, clinical practice, program improvement, assessment of student learning, and methods for teaching students with learning disabilities. She has published in journals such as *Intervention in School and Clinic* and *Learning Disability Quarterly.*

**Dr. Jennifer McGee** is Assistant Professor in the Department of Curriculum and Instruction. Her research interests include validity and measurement, program evaluation, self-efficacy, assessment, and STEM education. She has published in journals such as *the Journal of Psychoeducational Assessment, Action in Teacher Education,* and the *International Journal of Educational Research.*