RESEARCH & PRACTICE IN ASSESSMENT

RPA

2020 Virginia Assessment Group Annual Conference

NOVEMBER 11–13, 2020
Delta Hotels Marriott Richmond Downtown
www.virginiaassessment.org

CALL FOR PAPERS

Research & Practice in Assessment is currently soliciting articles and reviews for Volume 15 of the journal. Manuscripts submitted to RPA may be related to various higher education assessment themes, and should adopt either an assessment and measurement or an assessment policy/foundations framework. Contributions are accepted at any time and will be published as they are ready for publication. Manuscripts must comply with the RPA Submission Guidelines and be submitted to our online manuscript submission system found at www.rpajournal.com.

RESEARCH & PRACTICE IN ASSESSMENT

The goal of Research & Practice in Assessment is to serve the assessment community as an online journal focusing on higher education assessment. It is dedicated to the advancement of scholarly discussion amongst researchers and practitioners in this evolving field. The journal originated from the Board of the Virginia Assessment Group, one of the oldest continuing professional higher education assessment organizations in the United States. Research & Practice in Assessment is a peer-reviewed publication that uses a double-blind review process. Approximately forty percent of submissions are accepted for issues that are collated at least once annually. Research & Practice in Assessment is listed in Cabell's Directory and indexed by EBSCO, ERIC, Gale, and ProQuest.

TABLE OF CONTENTS

FROM THE EDITOR

Rolling Into the New Decade

"Without continual growth and progress, such words as improvement, achievement, and success have no meaning." – Benjamin Franklin

As we enter the new decade, the second issue of my tenure as Editor-in-Chief of Research & Practice in Assessment also stands as my last; at least in the traditional sense. Over the last decade, our journal has published one volume, split into two issues, per year. The issue-based format was largely adopted out of tradition; taking our lead from print-only journals with the need to compile multiple articles into larger sets to make the cost of printing reasonable. However, as an online-only journal, we have the opportunity and flexibility to be more reactive to the needs of our readers. Thus, beginning immediately following this issue, the journal of Research & Practice in Assessment will transition to an article-level publication model.

Using this new model, the journal will continue its rigorous and supportive process of double-blind peer review as well as our well-regarded publication process including copyediting and graphic design work. However, we will now publish articles when they are ready, as individual works, instead of waiting for the next issue date. Individual articles will then be collated at a later time into a larger collection. We believe that an article-level publication model will benefit both authors, who are looking for faster publication times, and readers, who are looking to stay apprised of the most up-to-date work in the field of assessment. We look forward to exploring this new format with you and welcome any feedback as we make this transition.

This issue of RPA includes six peer reviewed articles that address a variety of the biggest issues in our field. Drewes, Scantlebury, and Soslau share their work to assess the coteaching model from students' perspectives. Logli then discusses her work to expand on the important discussion around culturally responsive and inclusive assessment practices. Lyons and Polychronopoulos examine a novel method for exploring the connection between student engagement and value-added scores. Jenks and Green then revisit the important topic of faculty and staff perceptions of assessment using Q methodology. Blau and Blau next provide an overview of their attempt to lessen resistance to assessment through the use of humor. Finally, Song and Elftman lay out their process of developing surveys using a mixed-methods approach.

I hope this issue of Research & Practice in Assessment starts the rock moving downhill for you to spur even more exciting and important work in the decade to come.

Best Regards,

Nicholas Curtis

Nicholas Curtis
Editor-in-Chief, *Research & Practice in Assessment*

Abstract

Pupils are another group of stakeholders in teacher preparation because most programs include clinical experiences for students enrolled in teacher education degrees. There is an increasing number of teacher education programs in the United States that are using a coteaching model for the student teaching practicum. The aim of this study was to develop an instrument that would examine coteaching contextual features from pupils' perspectives. This research reports the design, validation, and implementation of a coteaching survey for pupils in classes where student teacher candidates along with their cooperating teachers have taught the class. Data were collected from over 7,000 students aged from 10 to 18 years. Exploratory factor analysis (EFA) was used to establish a 23-item instrument with three subscales. The sub-scales were Respectful and Caring Environment, Engagement and Motivation for Learning, and Behavior and Classroom Management. Cronbach reliability for the scales ranged from .857 for Respectful and Caring Environment, .837 for Engagement and Motivation for Learning, and .685 for Behavior and Classroom Management.

AUTHORS

Andrea Drewes
Rider University

Kathryn Scantlebury and
Elizabeth Soslau
University of Delaware

Listening to the Missing Voices:
Students' Perspectives on Coteaching

Over the past decade, a quickly growing number of teacher education programs have begun implementing coteaching as a model for student teaching because it better supports pupil learning and contributes to the collaborative professional development of both the teacher candidate and the clinical educator[1]. The teacher preparation field is swiftly building consensus for the use of coteaching approaches to develop novice teachers' expertise and to increase learning opportunities for pupils. Yet important voices are missing from the research literature to support these claims: pupils' voices.

Pupils spend many hours in classrooms and often their perception of the learning environment can differ from teachers. Pupils become "'experts' in knowing the 'ins and outs'" of their classroom (Bayne, 2012, p. 243). They can provide insights on the classroom climate and assess the learning environment (Fraser, 2001; Moos, 1973; Walberg & Haertel, 1980) to assist in identifying coteaching practices that need to be strengthened or reformed. For more than 40 years, other researchers have collected pupils' perspectives of their classroom climate, also known as the psychosocial learning environment. These pupil perspectives illuminate the benefits and informative challenges leading to improved learning opportunities for children through a more positive and effective learning environment (see Fraser, 2012a for a full review of other classroom environment instruments).

In the few coteaching studies that included capturing pupils' perspectives on their own learning environment the pupils were coteachers themselves aiding the instruction of other peer students to develop deeper content knowledge understandings (e.g. Schultze & Nilsson, 2018; Woodburn, 2010). Our search of the literature failed to unearth any research focused on pupils' perspectives of coteaching when the coteaching pair included a clinical educator and a teacher candidate with the pupils solely positioned as learners. While teacher education and assessment are two highlighted areas for the utilization of classroom environment instruments (Fraser, 2012a), there are currently no valid and reliable survey instruments aimed at collecting

CORRESPONDENCE

Email
adrewes@rider.edu

[1] Clinical educators are the mentor in-service teachers who host teacher candidates during student teaching field experiences; teacher candidates are the preservice student teachers pursuing licensure and/or certification; pupils, as we refer to them in this article, are the K–12 student learners in the classroom environment.

pupils' perspectives on their own cotaught learning environment and experiences. This highlights a critical gap in our understanding of coteaching as a collaborative teacher preparation approach and one that must be addressed due to the mounting pressure from teacher education accreditation agencies for these types of data.

This study acknowledges that coteaching is becoming a ubiquitous approach across teacher preparation programs, and as such, we as faculty teacher educators must attend to national accreditation agencies' continuous push for valid and reliable ways to collect data for ongoing program improvement (i.e. CAEP, 2015). Thus, we developed and validated a survey instrument that garnered over 7,000 pupils' perspectives on their own classroom environment where at least one teacher candidate and one clinical educator were using a coteaching model. Specifically, the instrument includes psychosocial learning environment scales which correspond to the classic framework of Moos's (1974) dimensions for the classification of human environments which include pupils' perceptions of Relationships, aspects of Personal Development, and elements of System Maintenance and System Change in the classroom. The new instrument provides educational researchers a valid and reliable tool to collect pupils' perceptions of their classroom, while also accounting for the interpersonal factors between and amongst coteachers (teacher candidate and clinical educator). These factors, which could be overlooked by outside observers, are closely captured by pupils positioned as classroom insiders (Fraser, 1998). In this way, pupils' voices are no longer missing in the data and can now be used to support coteaching model improvements, while also enabling faculty to attend to accreditation pressures for valid and reliable instruments.

Novel approaches to evauation and current interpretations of well-established scales still are forthcoming for educational research in practical applications

Theoretical Basis

Since the early 1980s, science education researchers have developed survey instruments that asked pupils for their perceptions of their classroom environment (Fraser, 2012a). Pupils spend many hours in classrooms developing their perceptions of the environment (Fraser, 2001), and as such, can be key informants. Researchers have used prior lines of research to ask pupils about the overall classroom climate and psychosocial elements, such as student-student and student-teacher relationships and how these relationships impact student engagement and learning behaviors (Bayne, 2012; Bell & Aldridge, 2014; Fraser, 1998).

Frequently these investigations and the resulting classroom environment instruments have widely relied on the theoretical backbone established by Moos's (1974) schema for identifying and describing a wide variety of environmental contexts for learning. Relationships are the focus of the first dimension in which researchers attend to the depth of support, respect, and involvement that teachers and students alike have with one another. The second dimension, Personal Development, highlights the opportunities for learning and growth provided by the environment. The final dimension, System Maintenance and System Change, describes the degree to which an environment is structured and orderly—as well as how, or if, it is responsive to change. Fraser (2012a) has extensively reviewed and described many frequently employed classroom environment scales according to these dimensions. For example, the Classroom Environment Scale (CES; Moos & Trickett, 1987) has scales related to Involvement, Affiliation, Teacher Support (all Relationship dimensions), Task Orientation, Competition (both Personal Development dimensions), Order and Organization, Role Clarity, Teacher Control, and Innovation (System Maintenance and System Change dimensions).

Yet novel approaches to evaluation and current interpretations of well-established scales still are forthcoming for educational research in practical applications. Bayne utilized the Constructivist Learning Environment Survey (CLES; Taylor, Fraser, & Fisher, 1997), reanalyzed the survey to classify four scales instead of five via quantitative analysis, and supplanted these findings through qualitative analysis (Bayne, 2012). New perspectives on this well-established body of knowledge are still needed to explore and understand emerging models for teaching and learning, especially from the pupil perspective. Namely, in this study we sought to better appreciate the psychosocial climate and contextual elements of the coteaching classroom through examination of the perceptions of the pupils experiencing such environments.

Coteaching was introduced as a model for student teaching in the early 2000s (Martin 2009; Murphy & Scantlebury, 2010; Roth, Tobin, Carambo, & Dalland, 2004; Tobin & Roth, 2006; Tobin, Zurbano, Ford, & Carambo, 2003). In the United States, teacher accreditation bodies such as the National Council for Accreditation of Teacher Education (NCATE, 2010), the Council for the Accreditation of Educator Preparation (CAEP, 2013), and the American Associate of Colleges of Teacher Education (AACTE, 2018) have noted the importance of clinical experiences in the education of a teacher and placed an increased emphasis on meaningful collaborations between stakeholders in teacher preparation programs, teachers, and schools. NCATE's Blue Ribbon panel praised coteaching as a model for linking theory and practice in preparing teachers to teach (NCATE, 2010). Likewise, AACTE's recent commission on clinical practice had coteaching as a key tenet in preparing teachers. Teacher education programs have adopted this model for student teaching because there is evidence that coteaching supports clinical educator and teacher candidate professional learning (Gallo-Fox & Scantlebury, 2015; Guise et al., 2016; Hedin & Conderman, 2015; Kerin & Murphy, 2015; Martin, 2009; Roth et al., 2004; Siry, 2011; Soslau, Gallo-Fox, & Scantlebury, 2018a) and student learning (Bacharach, Heck, & Dahlberg, 2010; Emdin, 2007). Recently programs have expanded the use of coteaching and have been using qualitative approaches to explore, to provide deeper descriptions of the approach, and to discern learning benefits to teacher candidates, pupils, and clinical educators (Gallo-Fox & Scantlebury, 2016; Milne, Scantlebury, Blonstein, & Gleason, 2011; Scantlebury, Gallo-Fox, & Wassell, 2008; Soslau et al., 2018a; Soslau, Kotch-Jester, Scantlebury, & Gleason, 2018b).

However, few studies have used quantitative methods to ascertain pupils' perspectives on the impact of coteaching on their classroom experiences. Murphy and Beggs (2010) reported primary school pupils' science attitudes improved after being cotaught for six months by a teacher candidate, majoring in science, in partnership with their classroom clinical educator (Murphy & Beggs, 2005; Murphy, Beggs, Carlisle, & Greenwood, 2004). In the United States, Bacharach, Heck and Dahlberg (2010) found pupils' reading and mathematics scores improved when they were cotaught. Though achievement outcomes are important, to understand how these outcomes were achieved we suggest that researchers should explore the classroom environment aspects of coteaching contexts. Quantitative data, which examines pupils' perspectives about the sociocultural climate of their cotaught classes, is needed to fully explore which features of coteaching may relate to improved pupil learning outcomes and attitudes. To date, these types of data have not yet been collected. The aim of this study was to develop instruments that could examine the coteaching contextual features from pupils' perspective.

Accreditation Requirements

In the United States, external teacher accreditation agencies, such as CAEP, review and evaluate teacher education programs. One criteria of the accreditation process is that programs must rely on "relevant, verifiable, representative, cumulative and actionable measures, and produce empirical evidence that interpretations of data are valid and consistent" (CAEP, 2015, para. 2). However, to date there has been no development of instruments that could collect valid and reliable data on pupils' and teachers' perceptions of the coteaching environment. The study research site, State University (a pseudonym) is required by its state department of education to complete CAEP accreditation. The university moved to develop valid and reliable survey instruments to document the perceptions and experiences of coteachers and pupils engaged in coteaching through the hosting of teacher candidates. In addition to satisfying accreditation requirements, another intended outcome of this research is to provide a valid instrument that can be used by other teacher preparation programs responsible for collecting data with valid and reliable instruments. Ultimately, our hope is that our survey instrument will be useful in evaluating coteaching as a model for student teaching across programs nationally and internationally.

Methods

Factor analysis examines patterns of variance and correlation (covariance) within participant responses on a survey instrument. Exploratory factor analysis (EFA) begins with all

Our hope is that our survey instrument will be useful in evaluating coteaching as a model for student teaching across programs nationally and internationally

items and uncovers related latent variables. These items are then grouped into subsets based on participants' patterns of responses. The main goal of EFA is to identify these sets of items and does not base the organization of survey items to any a priori theory.

The survey development process was influenced by the meta-framework presented by Onwuegbuzie, Bustamante, and Nelson (2010) for a mixed methods development process and the four-step procedure established for developing and validating measures (Crocker & Algina, 1986; Sax 1997). These frameworks guided our approach to the creation of possible survey items, testing, and refinement of these items, all while utilizing both qualitative and quantitative approaches to develop a measure that would allow the pupils to share their insights and perceptions of the coteaching classroom environment. While taking this approach, we worked to develop the survey to operationalize the practices and outcomes that we could expect these pupil stakeholders to experience across a variety of coteaching settings.

Procedure

This study was completed within a larger ongoing research study on the impact of coteaching on clinical educators, teacher candidates, and their pupils. This pilot study was conducted under the established IRB protocols. Participants were recruited from classes cotaught by teacher candidates from grades 4 (pupils aged around 10 years) through high school (pupils aged up to 18 years old). Due to limited technology access at some schools, pupils completed a pen and paper version of the survey to maximize participation.

In the initial item development, our research team qualitatively reviewed existing surveys related to student teaching already in use by State University and other similarly purposed surveys from other institutions. These surveys provided a preliminary framework for designing appropriate demographic and background questions to allow the instrument to be employed in diverse coteaching settings.

In the next phase of item development, the lead author collected and reviewed numerous coteaching-related articles through a literature review to create an initial list of survey items. All survey items were structured as Likert type responses for evaluation of frequency with a 5-point scale (i.e. "Never," "Mostly Not," "Sometimes," "Mostly Yes," "Always") with an unclear option of "I Don't Know."

After an initial pilot in Spring 2015, which garnered 341 responses, the survey consisted of 28 original items. At this point the research team (comprised of university-based field instructors, faculty experienced with the coteaching model, and clinical educators) qualitatively reviewed the survey utilizing a thematic analysis approach to evaluate items for clarity and to establish face validity and content validity based on their collective expertise in this research area. Three scales were preliminarily identified through collaboration and consensus building with researchers and clinical educators.

During this qualitative review process five items were dropped and six items were revised. Of the items that were revised, changes were made to improve interpretation by removing double barrel items (i.e. old version: "If I have questions or need something, I can go to any of my teachers" revised to new version "If I need something, I can go to any of my teachers") and/ or to allow students to be unbiased in their responses (i.e. old version: "If I break the rules, both teachers would give me the same punishment" revised to "If a student breaks the rules, both teachers would give the same consequence"). Several items were removed, such as one item related to technology usage. These eliminations were designed to focus the survey on research themes and outcomes identified by prior research in the coteaching classroom and to be mindful of the negative impact on responses due to a lengthy survey (Galesic & Bosnjak, 2009; Schwarz, Groves, & Schuman, 1998). Following these revisions, additional testing of the 23-item survey was completed at the end of the Fall 2015, Spring 2016, and Fall 2016 semesters. No further edits were made to the items during this data collection period. During the testing cycles over 7,620 surveys were collected from pupils.

All survey items were structured as Likert type responses for evaluation of frequency with a 5-point scale

The process to validate the survey relied on Principal Component Analysis, an EFA technique, with varimax rotation due to the orthogonal nature of the data. This type of factor analysis relies on the judgement of the researcher along with the analytic results to determine the appropriate number of factors or components to include (Tabachnick & Fidell, 2007). The most common method is to retain all factors with an eigenvalue over 1.0, but often this method will over- or under-estimate the true number of dimensional factors (Gorsuch, 1983; Zwick & Velicer, 1986). Parallel analysis is one of the best methods for ascertaining the appropriate number of components and scree plot examination is a useful supplemental method (Velicer, Eaton, & Fava, 2000; Zwick & Velicer, 1986). After utilizing all three approaches we determined that three components were present. Then the research team evaluated each group of items to characterize them by their similarities and to name each component, herein referred to as the survey scales.

Interpretation of the EFA pattern matrix was conducted employing the guidelines for factor loadings where greater than .71 is excellent, greater than .63 is very good, greater than .55 is good, and greater than .45 is fair (Comrey & Lee, 1992). Typically items with factor loadings less than .4 are removed or revised, though they may be kept under special circumstances—especially when sample sizes are larger (Field, 2013).

Reliability for each scale was considered by a Cronbach's alpha analysis. Desired values for Cronbach's alpha are generally between .7 and .8, while greater than .9 indicates an overly homogenous or repetitive grouping, between .5 and .7 is deemed acceptable, and lower than .5 is considered unacceptable (George & Mallery, 2003).

After the survey scales were established and evaluated for reliability, another analysis was conducted via a one-way ANOVA to determine if the different pupil groups were responding differentially to the survey. Lastly, a descriptive analysis for the individual item means and standard deviations was completed for the entire sample population to initiate an investigation into the overall findings uncovered by the survey. Prominent findings from the most and least positively rated items will be examined further in the results and discussion sections.

Participants

The participants in this study were pupils from classrooms in which coteaching was used as a model for the student teaching experience. Student teaching consisted of at least 12 and no more than 15 weeks, full time, with daily participation in the school setting. Teacher candidates were involved with all aspects of planning, implementing and evaluating the curriculum, in collaboration with at least one experienced clinical educator. The pupils were from elementary classrooms (grades 4 to 5, ages 9 to 11 years, N = 1,333), middle school classrooms (grades 6 to 8, ages 11 to 13 years, N = 6,059), and high school classrooms (grades 9 to 12, ages 14 to 17 years, N = 228). The total population of pupils, from which the sample was selected, was majority (> 60%) white and middle class with approximately half of the population identifying as female and the other half as male.

Results

Factor Analysis Results

The final survey consists of three scales: Respectful & Caring Environment (9 items), Engagement & Motivation for Learning (8 items), Behavior & Classroom Management (6 items). Multiple analytic methods (scree, eigenvalues, and parallel analysis) converged such that three scales were present in the data. The factor analysis model appropriately fit the data (determinant = .001, KMO = .930, Bartlett's < .001). The final survey consists of 23 items and exemplar items for each of the three scales are presented in the appendix. For a full text copy of the survey with all survey items, please contact the corresponding author of this article.

Scale reliability as measured by Cronbach's alpha was .857 for Respectful & Caring Environment, .837 for Engagement & Motivation for Learning, and .685 for Behavior &

The final survey consists of three scales: Respectful & Caring Environment (9 items), Engagement & Motivation for Learning (8 items), Behavior & Classroom Management (6 items)

Classroom Management. Table 1 provides the factor loadings by scale. Note that questions 3, 5, and 20 are reverse-coded items.

Table 1
Factor Loadings by Subscale

	Subscale		
	Respectful & Caring Environment	Engagement & Motivation for Learning	Behavior & Classroom Management
Q11	.746		
Q7	.711		
Q13	.688		
Q9	.679		
Q15	.640		
Q4	.608		
Q16	.608		
Q6	.559		
Q17	.554		
Q22		.781	
Q21		.738	
Q18		.715	
Q19		.680	
Q12		.636	
Q23		.616	
Q14		.565	
Q1		.425	
RC Q3			.649
Q8			.606
RC Q20			.570
Q2			.522
RC Q5			.415
Q10			.282

Item Q10 ("It is clear who is in charge of the classroom.") has a low factor loading of .282 on Scale 3—Behavior and Classroom Management. Due to the large sample size, there was both quantitative support, and a qualitative necessity, for retaining this item. After discussion with the research team, it was determined that this item could prove qualitatively insightful for current and future coteaching research studies because shared power and collaboration are hallmarks of quality coteaching (Scantlebury et al., 2008; Soslau et al., 2018a, 2018b). Therefore, the item was retained despite a low factor loading.

ANOVA results

The one-way ANOVA was completed utilizing the survey responses for the entire survey to compare how the three different groups of pupils responded to the questions. Levene's test is significant, which indicates a violation of the homogeneity of variance assumption (this is likely due to the drastically different group sizes). Therefore, the Welch statistic from the Robust Tests of Equality of Means is reported instead of the traditional ANOVA F statistic. The results indicate there was a statistically significant difference in how the three groups are responding to the survey (Welch statistic = 38.374; df1 = 2; df2 = 565.053; sig < .001).

Post hoc comparisons

Games-Howell post hoc test was utilized due to the violation of the homogeneity of variance assumption. Table 2 presents the overall survey means and standard deviations for each group of pupils as well as the sample size for each group.

Table 2
Survey Means, SD, and N by School Level

Group	Mean	SD	N
Elementary	4.14	.48	1331
Middle	4.05	.54	6056
High	3.78	.67	228

Table 3 shows the results of the Games-Howell post hoc test. There was a significant small effect between elementary and middle school pupils (Cohen's d = 0.180), and a significant medium effect between elementary and high school pupils (Cohen's d = 0.673), and middle and high school pupils (Cohen's d = 0.459).

Table 3
Post hoc results from Games-Howell analysis

Groups compared	*p*-value	Cohen's *d*	Effect
Elem vs Middle	<.001	.180	Small
Elem vs High	<.001	.673	Medium
Middle vs High	<.001	.459	Medium

This post hoc analysis demonstrates that not only is the difference between the three groups statistically significant, but that elementary pupils are answering the survey most positively (M = 4.14) and high school pupils less positively (M=3.80) with middle school pupils between these two groups (M=4.05). The difference between elementary and middle school pupils represents a small effect size, while the differences between elementary and high school as well as middle and high school pupils demonstrates a medium effect size. With further data collection to gain group sizes that are more equivalent, these results may change, and this will be considered in future analysis. However, because only small to medium effect sizes are observed, we consider this survey to be appropriate for use across the grade bands tested thus far (grades 4 to 12, pupils aged 10 to 18 years).

Elementary pupils are answering the survey most positively (M = 4.14) and high school pupils less positively (M=3.80) with middle school pupils between these two groups (M=4.05)

Descriptive Analysis for Scales and Individual Survey Items

Table 4 presents the means and standard deviations for each of the three scales. The scale, Respectful & Caring Environment, had the highest mean score of 4.57 of the three scales, with Engagement & Motivation for Learning having the lowest mean score of 3.52.

Table 4
Mean and Standard Deviation by Subscale

	Mean	SD	N
Respectful & Caring Environment	4.57	.546	7620
Engagement & Motivation for Learning	3.52	.796	7620
Behavior & Classroom Management	3.86	.779	7620

Table 5 is the descriptive analysis results for each item individually. The three questions currently highlighted yielded the lowest mean responses (least positive); the four questions emphasized with an underline garnered the highest mean responses (most positive).

Table 5
Mean and Standard Deviation by Item

Item	Mean	SD	Item	Mean	SD
Q1	3.65	.907	Q13	4.60	.713
Q2	3.82	1.144	Q14	3.67	1.169
Reverse code Q3	3.12	1.309	Q15	4.66	.665
Q4	4.79	.600	Q16	4.38	.876
Reverse code Q5	4.31	1.094	Q17	4.33	.925
Q6	4.66	.711	Q18	3.77	1.139
Q7	4.78	.558	Q19	3.41	1.158
Q8	4.28	.895	Reverse code Q20	3.94	1.301
Q9	4.47	.836	Q21	3.24	1.274
Q10	4.33	.962	Q22	3.39	1.223
Q11	4.66	.700	Q23	3.43	1.161
Q12	3.81	1.110			

This survey can be used to support program evaluation for improvement and for reporting to accreditation agencies

After completing data collection, and conducting scree, eigenvalues, and parallel analysis, three reliable scales emerged: Respectful & Caring Environment (9 items); Engagement & Motivation for Learning (8 items), and Behavior & Classroom Management (6 items). Pupils' responses indicated a violation of the homogeneity of variance assumption possibly due to the unequal sample size that reflect the lower numbers of secondary science student teachers compared with those student teachers preparing to teach elementary and/or middle school classes. Clear patterns emerged with the scales and individual items. Questions 3, 21, and 22 had the lowest mean scores of the individual items, while questions 4, 6, 7, and 11 had the highest mean scores of the individual items

Discussion

There are two main findings that extend the current body of coteaching research in practice. First, this survey instrument is a valid and reliable way to gather pupils' perspectives about their classroom contexts for learning within a coteaching model for student teaching. These collected perceptions are quite valuable as they relate to student behavior and other achievement outcomes and cannot be omitted when determining instructional effectiveness (Fraser, 2012b). The use of student-oriented classroom environment surveys has meaningful application for the purpose of research and evaluation of educational innovations (Fraser, 2012a). Additionally, this survey can be used to support program evaluation for improvement and for reporting to accreditation agencies, which represent two other meaningful applications espoused by Fraser (2012a) for teacher education and teacher assessment.

Heretofore, only teachers' voices have been included when evaluating the implementation of coteaching. Yet pupil voices must be influential in our work as educators because pupils "have the potential to assess and evaluate honestly and flawlessly" (Issa, 2015, p. 106) the classroom environment, especially the psychosocial or relational aspects within the classroom climate. These current data provide an opportunity for researchers to triangulate findings with teacher data to make more robust attributions about the quality of coteaching in classrooms and within teacher preparation programs. Teacher candidates and clinical educators can also use the survey results to reflect upon how their coteaching influenced classroom learning, especially along the dimensions of Moos's framework: Relationships, Personal Development, and System Maintenance and System Change (1974). Many clinical educators regularly support teacher education programs by mentoring teacher candidates and by having access to pupils' perspectives on their learning experiences; this may result in clinical educators and teacher education programs collaboratively discussing changes to improve field experiences more broadly.

We also provide useful exemplars of findings from the survey with an eye to how this survey moves the field of teacher education research and assessment to consider the impact of coteaching models. The three scales, Respectful and Caring Environment, Engagement and Motivation for Learning, and Behavior and Classroom Management are reliable scales that are aligned with the theory and purpose of using coteaching as a model for student teaching. Items on the Respectful and Caring Environment scale ask pupils their perception of whether the coteachers worked well together and if they respected each teacher. Scantlebury et al. (2008) noted that successful coteaching between teacher candidates and clinical educators require corespect as a prerequisite. If coteachers did not respect each other then typically they would fail in implementing the most basic tenets of coteaching. In this scale, we see counterparts to the Relationship Dimension from Moos (1974) and other traditional survey instruments. For example, a representative item from the Classroom Environment Scale (CES),"The teacher takes a personal interest in the students," closely parallels an important assumption for coteaching that coteachers are focused on pupils' learning. Other items on this scale for a Respectful and Caring Environment include statements such as "All my teachers make me feel like I can learn" and "I remember the lesson better when I have more than one teacher." These items work to examine whether the additional human resources in the classroom (that is, the teacher candidate) support pupils' perceptions that their learning needs are being addressed.

The second scale of this new instrument relates to Engagement and Motivation for Learning. Again, the traditional theoretical framework put forth by Moos holds here as well as similarities are evident to the Personal Development dimension. In our previous research, coteachers reported that the model provided new opportunities for teachers to implement new or different pedagogical approaches (Gallo-Fox & Scantlebury, 2015). Items such as "When there is more than one teacher, we do more activities in class" and "I participate more often when I have more than one teacher" may help to strengthen coteachers' self-reports, which claim that coteaching served to expand their pedagogical repertoire during student teaching—increases reported both by the teacher candidate and the clinical educator (Soslau et al., 2018a).

The third scale, Behavior and Classroom Management, includes items such as "Sometimes I ask one teacher instead of the other teacher for permission because I know they will let me do what I want to do," which seeks to unearth pupils' perceptions about how coteachers manage the class and if they perceive that coteachers share this responsibility. Once more, this final scale complements the third dimension of System Maintenance and System Change (Moos, 1974). An item from the CES is representative of this third dimension: "There is a clear set of rules for students to follow." Items from our Behavior and Classroom Management scale pivot to address the unique circumstances of multiple teachers in the cotaught classroom (i.e. "If a student breaks a rule, both teachers would give the same consequence" and "Each teacher has different rules"). However, our analysis continues to support and confirm the continuing utility of these three dimensions of the learning environment.

Our analysis continues to support and confirm the continuing utility of these three dimensions of the learning environment

Additionally, we know that classroom management is a primary focus of coteachers' huddles. Huddles are the impromptu meetings between teacher candidates and classroom teachers that often occur during instruction to check in about how the lesson is unfolding and discuss any necessary changes (Soslau et al., 2018b). Coteachers must set clear expectations for instruction, and through the use of huddles, instruction can be nimble and responsive to the needs of the classroom environment and the students.

We also know that teacher candidates report finding their participation in these help-seeking huddles educative, thus coteaching functions as a context for learning how to improve their management skills (Soslau et al., 2018b). This scale looks to ascertain pupils' perceptions of the shared responsibility for supporting pupils' behavior. This, once again, provides additional data that can be used to bolster claims about the teacher candidates' self-reports regarding the usefulness of their participation in huddles focused on classroom management.

In addition to serving as an extension to the existing research on coteaching, there are practical implications that are critical for us as teacher educators in the local context. The scales on the pupil survey will provide data to revise and strengthen the professional development towards the goal of improving coteachers' learning opportunities for all pupils when using coteaching as a model for student teaching. Bayne describes this practical application for all: "Learning more about how the learning environment is experienced has potential for creating more adaptable forms of teaching, learning and assessing—including assessing the learning environment itself— that can support a diversity of students" (2012, p. 246).

Researcher-practitioners may consider collecting and sharing survey data with the specific clinical educators and teacher candidates that serviced the students. In this way, survey data can serve as a formative assessment for the coteaching pair to reflect on not only their enactment of coteaching practices, but also how those coteaching practices were experienced by the pupils. This could work to ensure that the classroom climate is a comfortable and productive space (Aldridge et al., 2016). Methods faculty, who teach teacher candidates before the student teaching practicum, may also find these pupil data compelling because part of these faculty's charge is to support teacher candidates' development of collaborative expertise—a foundational competency for coteachers. By prioritizing data from pupils' perspectives, instruments like the one developed in this study can offer a more detailed picture of the coteaching process (Fraser, 1998). Moreover, while the survey was developed to address pupils' perceptions of coteaching with teacher candidates involved in the teaching process, the items and scales are also applicable to other coteaching settings, such as when a special education teacher coteaches with a general education teacher. The survey items may also be useful for teachers who coteach with pupils, parents, or other volunteers with the goal of supporting their pupils' academic and social-emotional wellbeing as key indicators of a positive psychosocial learning environment and along Moos's three dimensions of the learning environment.

The survey items may also be useful for teachers who coteach with pupils, parents, or other volunteers

Limitations of this work emerged due to the unequal pupil sample sizes, a consequence of the varying enrollment in State University's teacher education programs. Future research is underway to link pupil survey responses with survey responses from the clinical educators and teacher candidates to further develop a more detailed research imagery of the coteaching classroom in practice. Overall, this study shows that pupils are sensitive to the differences in a coteaching learning environment. Pupils can provide useful insight into the overall nature of the care and respect exhibited, the encouragement they encounter toward their learning, and the managerial approaches that are in place within a coteaching classroom.

Appendix.
Exemplar survey items for each scale

Scale 1 – Respectful & Caring Environment

Q11 - My teachers in this class care about me and my learning.

Q13 - My teachers teach well together.

Scale 2 – Engagement & Motivation for Learning

Q18 - I learn better when I have more than one teacher.

Q19 - When there is more than one teacher, we do more activities in class.

Scale 3 – Behavior & Classroom Management

Q2 - If a student breaks the rules, both teachers would give the same consquences.

Q8 - I know that both teachers use the same rules for students.

References

Aldridge, J. M., Fraser, B. J., Fozdar, F., Ala'i, K., Earnest, J., & Afari, E. (2016). Students' perceptions of school climate as determinants of wellbeing, resilience and identity. Improving Schools, 19(1), 5–26.

American Associate of Colleges of Teacher Education (AACTE) (2018). A pivot towards clinical practice, and the renewal of educator preparation. Washington, DC.

Bacharach, N., Heck, T. W., & Dahlberg, K. (2010). Changing the face of student teaching through co-teaching. Action in Teacher Education, 32(1), 3–14. doi: 10.1080/01626620.2010.10463538.

Bayne, G. U. (2012). Capturing essential understandings of the urban science learning environment. Learning Environments Research, 15(2), 231–250.

Bell, L. M., & Aldridge, J. M. (2014). Student voice, teacher action research and classroom improvement. Rotterdam, Netherlands: Sense.

Comrey, A. L., & Lee, H. B. (1992). A first course in factor analysis (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Council for the Accreditation of Educator Preparation (CAEP) (2015). The CAEP Standards. Retrieved from http://caepnet.org/standards/introduction.

Crocker, L. M., & Algina, J. (1986). Introduction to classical and modern test theory. New York, NY: Holt, Rinehart and Winston.

Emdin, C. (2007). Exploring the contexts of urban science classrooms: Part 2. The role of rituals in communal practice. Cultural Studies of Science Education (2), 351–373. doi:10.1007/s11422-007-9057-x.

Field, A. (2013). Discovering statistics using IBM SPSS statistics. Thousand Oaks, CA: SAGE

Fraser, B. J. (1998). Science learning environments: Assessment, effects and determinants. In B. J. Fraser & K. G. Tobin (Eds.), The international handbook of science education (pp. 527–564). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Fraser, B. J. (2001). Twenty thousand hours: Editor's introduction. Learning Environments Research, 4(1), 1–5.

Fraser, B. J. (2012a). Classroom environment instruments: Development, validity and applications. Learning Environments Research, 1(1), 7–33.

Fraser. B. (2012b). Classroom learning environments: Retrospect, content and prospect. In B. Fraser, K. Tobin, & C. McRobbie (Eds.) Second International Handbook of Science Education. (pp. 1191-1239). Dordrecht, Netherlands: Springer.

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey, Public Opinion Quarterly, 73(2), 349–360. doi:0.1093/poq/nfp031

Gallo-Fox, J., & Scantlebury, K. (2015). ""It isn't necessarily sunshine and daisies every time:" Coplanning opportunities and challenges when student teaching. Asia-Pacific Journal of Teacher Education. doi:10.1080/1359866X.2015.1060294

Gallo-Fox, J., & Scantlebury, K. (2016). Coteaching as professional development for cooperating teachers. Teaching and Teacher Education. E doi:10.1016/j.tate.2016.08.007

George, D., & Mallery, P. (2003). SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th ed.). Boston: Allyn & Bacon.

Gorsuch, R. L. (1983). Factor Analysis (2nd Ed.). Hillsdale, NJ: Erlbaum.

Guise, M., Habib, M., Robbins, A., Hegg, S., Hoellwarth, C., & Stauch, N. (2016). Preconditions for success and barriers to implementation. Teacher Education Quarterly, 43(4), 55–76. doi:10.1016/j.tate.2017.05.002.

Hedin, L., & Conderman, G. (2015). Shared promises and challenges of coteaching: general-special education and mentor preservice partnerships. Action in Teacher Education, 37(4), 397–417. doi:10.1080/01626620.2015.1078756.

Issa, H. M. (2015). The effectiveness of faculty members in managing and arranging psycho-social environment at Al Ain University of Science and Technology, International Journal of Humanities and Social Science, 5(1), 98–107.

Kerin, M., & Murphy C. (2015). Exploring the impact of coteaching on pre-service music teachers. Asia-Pacific Journal of Teacher Education, 10, 1–17.

Martin, S. (2009). Learning to teach science. In K. Tobin & W.-M. Roth (Eds.), World of science education: North America (pp. 567–586). Netherlands: Sense Publishers.

Milne, C., Scantlebury, K., Blonstein, J. & Gleason, S. (2011). Coteaching and disturbances: Building a better system for learning to teach science. Research in Science Education, 41, 414–440. doi:10.1007/s11165-010-9172-7

Moos, R. H. (1973). Conceptualization of Human Environments. American Psychologist, 28(8), 652–665.

Moos, R. H. (1974). The social climate scales: An overview. Palo Alto, CA: Consulting Psychologists Press.

Moos, R.H. & Trickett, E.J. (1987). Classroom Environment Scale manual (2nd ed.). Palo Alto, CA: Consulting Psychologists Press.

Murphy, C., & Beggs, J. (2005) Co-teaching as an approach to enhance science learning and teaching in primary schools. In W-M Roth & K. Tobin (Eds.), Teaching together, learning together (pp. 207–231). New York: Peter Lang.

Murphy, C., & Beggs, J. (2010). A five-year systematic study of coteaching science in 120 primary schools. In C.

Murphy & K. Scantlebury (Eds.), Coteaching in international contexts: Moving forward and broadening perspectives, (pp. 11–34). New York, NY: Springer. doi:10.1007/978-90-481-3707-7_2

Murphy, C., Beggs, J., Carlisle, K., & Greenwood, J. (2004). Students as 'catalysts' in the classroom: The impact of co-teaching between science student teachers and primary classroom teachers on children's enjoyment and learning of science. International Journal of Science Education, 26(8), 1023–1035. doi:10.1080/1468181032000158381

Murphy, C., & Scantlebury, K. (Editors). (2010). Coteaching in international contexts: Research and practice. London: Springer Publishing.

National Council for Accreditation of Teacher Education (NCATE) (2010). Transforming teacher education through clinical practice: A national strategy to prepare effective teachers. Washington, DC.

Onwuegbuzie, A. J., Bustamante, R. M., & Nelson, J. A. (2009). Mixed research as a tool for developing quantitative instruments. Journal of Mixed Methods Research, 4(1), 56–78. doi:10.1177/1558689809355805

Roth, W.-M., Tobin, K., Carambo, C., & Dalland, C. (2004). Coteaching: Creating resources for learning and learning to teach chemistry in urban high schools. Journal of Research in Science Teaching, 41(9), 882–904. doi:10.1002/tea.20030

Sax, G. (1997). Principles of educational and psychological measurement and evaluation. Belmont, CA: Wadsworth.

Scantlebury, K. Gallo-Fox, J. & Wassell, B. (2008). Coteaching as a model for preservice secondary science teacher education. Teaching & Teacher Education, 24, 967–981. doi.org/10.1016/j.tate.2007.10.008

Schultze, F., & Nilsson, P. (2018). Coteaching with senior students – a way to refine teachers' PCK for teaching chemical bonding in upper secondary school. International Journal of Science Education, doi:10.1080/09500693.2018.1436792

Schwarz, N., Groves, R.M., & Schuman, H. (1998). Survey methods. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), The handbook of social psychology (4th ed., Vol. 2; pp. 143–179). New York, NY: McGraw-Hill.

Siry, C. A. (2011). Emphasizing collaborative practices in learning to teach: Coteaching and cogenerative dialogue in a field-based methods course. Teaching Education, 22(1), 91–101. doi:10.1080/10476210.2010.520699.

Soslau, E., Gallo-Fox, J., & Scantlebury, K. (2018a). The promises and realities of implementing a coteaching model of student teaching. The Journal of Teacher Education. doi:10.1177/0022487117750126

Soslau, E., Kotch-Jester, S. Scantlebury, K., & Gleason, S. (2018b). Coteachers' huddles: Developing adaptive teaching expertise during student teaching. Teaching and Teacher Education, 73, 99–108, doi:10.1016/j.tate.2018.03.016

Tabachnick, B. G., & Fidell, L. S. (2007). Using multivariate statistics (5th ed.). Boston, MA: Pearson.

Taylor, P. C., Fraser, B. J., & Fisher, D. L. (1997). Monitoring constructivist classroom learning environments. International Journal of Educational Research, 27, 293–302.

Tobin, K., & Roth, W.-M. (2006). Teaching to learn: A view from the field. Rotterdam: Sense.

Tobin, K., Zurbano, R., Ford, A., & Carambo, C. (2003). Learning to teach through coteaching and cogenerative dialogue. Cybernetics and Human Knowing, 10(2), 51–73.

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helms (Eds.), Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy. (pp. 41–71). New York, NY: Guilford.

Walberg, H. J., & Haertel, G. D. (1980). Validity and use of educational environment assessments. Studies in Educational Evaluation, 6(3), 225–238.

Woodburn, C. A. (2010). Students as coteachers in an urban high school mathematics class. In C. Murphy, & K. Scantlebury (Eds.), Coteaching in international contexts: Research and practice (pp. 369–381). Dordrecht, Netherlands: Springer.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. Psychological Bulletin, 99, 432–442.

Abstract

This paper argues for organic initiatives that supplement the assessment work already being done, by blooming naturally from faculty and staff's expertise while also pollinating each dimension of assessment with cultural responsiveness. The theoretical framework draws from two leading paradigms—Culturally Responsive Assessment and Assessment 2.0. The Grounded Theory methodology uses Universal Design for Learning (UDL) guidelines to analyze (a) course-level assessment reports, (b) campus-wide activity on student engagement, and (c) campus-wide discussions around assessment, including gains and gaps in student learning. The study finds that assessment data emerge spontaneously across numerous spaces on campus, can be collected in flexible manners, and can be analyzed through an equity lens in order to support a diverse student population. Through this fresh approach findings show that faculty members are implementing inclusive practices in their instruction and assessment, and closely collaborate with units in student affairs to engage all learners.

AUTHORS

Chiara Logli

Honolulu Community College

Culturally Responsive Assessment 2.0: Revisiting the Quest for Equity and Quality in Student Learning

This paper examines how assessment can illuminate the extent to which teaching practices support learning of all students in a diverse environment. It highlights the significance of diversity theories and diversified methodologies in the field of assessment. On one hand, culturally responsive pedagogies have gained vast ground in education, but institutional assessment has only recently started to consider its intersect with equity (Montenegro & Jankowski, 2017b). While the process through which students demonstrate their knowledge and development is increasingly different, the criteria on which they are evaluated generally remain the same (Montenegro & Jankowski, 2017a). On the other hand, graduation rates, achievement gaps, institutional benchmarks, and other numerical data are insufficient to support equitable imperatives (Montenegro & Jankowski, 2017b). Assessment has to emerge organically from data sources that already exist, including thick and rich descriptions, beyond the preconceived evaluation plans and rationalized systems that still dominate literature and practice (Metzler & Kurz, 2018).

CORRESPONDENCE

Email
logli@hawaii.edu

The theoretical framework draws from the field of quality and equity assurance in higher education. However, both areas are often heavy in practical prescriptions and light in theoretical bases, as both diversity and assessment professionals "remain focused on the details of practice—getting it done" (Hershock, 2010; Metzler & Kurz, 2018, p. 4). Therefore, I build on relevant theorists inside and outside the field of assessment to contribute to a Culturally Responsive Assessment 2.0 paradigm. Culturally responsive assessment concerns matters of equity, such as shaping evaluation tools through culture-based lenses, disaggregating the data by student background, and using results to improve learning of all students (Montenegro & Jankowski, 2017b). Assessment 2.0 refers to flexible and yet robust approaches, including bottom-up processes, collective meaning-making, and organic assessment designs—whether formative, summative, quantitative, or qualitative (Metzler & Kurz, 2018). In this study, I integrate the two models because I claim that on one hand innovating approaches to assessment in the 21st century requires explicit culture-based considerations and on the other hand culturally relevant assessment can only exist within malleable paradigms as Assessment 2.0.

This pilot study represents a first attempt to bring a variety of organic data sources about student learning in conversation with one another through the lens of diversity. It triangulates three sets of data: (a) course-level assessment reports, (b) a campus-wide activity on student engagement, and (c) discussions around assessment at the Honolulu Community College in Hawaii (United States of America). The first instrument was designed as an assessment tool but provided wide flexibility, whereas the other two methods emerged organically outside preconceived assessment plans. The methodology follows Grounded Theory principles, including two cycles of coding and tallying via NVIVO software (Charmaz, 2010). Through the first coding cycle all emerging patterns resonated with the Universal Design for Learning (UDL) model, which is pertinent to our diverse and inclusive institution (Rao, in press). Therefore, the second coding cycle mapped the three datasets to UDL guidelines.

This paper argues for organic initiatives that supplement the assessment work already being done, by blooming naturally from faculty and staff's expertise while also pollinating each dimension of assessment with cultural responsiveness (Metzler & Kurz, 2018, p. 4; Montenegro & Jankowski, 2017b). The study finds that assessment data emerge spontaneously across numerous spaces on campus, can be collected in flexible manners, and can be analyzed through an equity lens in order to support a diverse student population. Through this fresh approach, findings show that faculty members are implementing inclusive practices in their instruction and assessment, and closely collaborate with units in student affairs to engage all learners.

This pilot study represents a first attempt to bring a variety of organic data sources about student learning in conversation with one another through the lens of diversity

Methodology: Grounded Theory Meets Universal Design for Learning

This study took place in 2018 at the Honolulu Community College in Hawaii—a very diverse campus in a very diverse state, with widespread commitment to indigeneity, inclusion, and equity (Accreditation Taskforce, 2018). It offers both vocational and liberal arts degrees, allowing students to transfer credits to achieve junior class standing at four-year higher education institutions within the state. It also offers noncredit courses in apprenticeship and continuing education as well as college credits for seniors in high school. Students at the college seek affordable and flexible education, as many have families, work full time, and return to college or the workforce after years of absence.

People of color constitute 91% of the students, including 47% Asian, 24% Hawaiian, and 13% mixed ethnicities (Arbuckle, 2018). The five pillars of the university highlight student focus, Hawaiian values, diversity, sustainability, as well as community partnerships. For instance, the Hulili Ke Kukui Hawaiian Center supports the enrollment pathway for Hawaiian students as well as the integration of Hawaiian language, culture, and values into the campus community, including through faculty and staff development (Hola Hou). A variety of platforms connect academic and student affairs in order to promote student retention and completion.

A triangulation approach is at the foundation of this study because a robust assessment program draws on multiple sources of evidence at multiple levels within the institution (Metzler & Kurz, 2018). The three data sources merged organically from improvement-based initiatives, but sat in compartmentalized spaces. This study is an attempt to break their isolation so they can talk with one another and create meaning together. The focus is on data sources that illustrate the paradigm shift from teaching to learning, which is at the core of our student-centered institution (Barr & Tagg, 1995). Meaningful assessment emerges when faculty and staff integrate assessment into their normal work (Allen, 2004). For instance, assignments play a key role in assessment because generally (a) faculty are strategically positioned to determine the quality of student work (Cain, 2014); in addition (b) students generally make their greatest efforts under required tasks, (c) assignments are pedagogically powerful in communicating faculty's expectations, and (d) making a difference in student learning happens in the day-to-day work (Hutchings, Jankowski, & Ewell, 2014). From classroom data practice can grow to departmental and institutional levels (Barr & Tagg, 1995).

First, I analyzed course-level assessment reports for the 2014–2018 cycle. These documents are preconceived assessment plans but leave room for personalized approaches as they ask broad questions about Course Learning Outcomes (CLOs). I examined CLO reports from three of the six campus academic divisions. The three divisions provided CLO reports for 210 out of their 267 courses (79%), covering 1833 CLO and including qualitative data analysis about assessment strategies that supported student mastery of the CLOs. Second, I examined index cards from a campus-wide activity that took place at commencement in Fall 2018, when a top administrator asked the staff and faculty members in attendance to write down their current practices for purposefully engaging students. One hundred and forty-five participants submitted 233 practices. This activity was not designed as an assessment plan and emerged organically.

Third, I analyzed notes that I took at campus-wide meetings, namely Assessment Taskforce, Assessment Townhall, Assessment Showcase, and assessment workshops as well as commencement and faculty development series, which featured faculty discussing best practices. These dialogues were not preconceived assessment activities per se; instead I saw them as glimpses into the many faculty conversations that informally take shape across campus about improving student learning and institutional performance. Capturing them was a challenge that was worth facing, since they are rich examples of thick description (Geertz, 1973). The dialogues engaged faculty and staff while fostering their ownership of assessment, as they are the creative forces in the process (Baker, Jankowski, Provezis, & Kinzie, 2012). These discussions provided the opportunity to analyze the course-level assessment reports and index cards. They became a space to validate the findings from the previous two methods. According to the participants, the results looked like "a typical day in the classroom." These campus-wide experiences were possible with the support of the administration, which allocated the necessary time, space, and resources while also endorsing the overarching flexible and organic approach to assessment on our campus (Baker et al., 2012).

My methodology follows a constructivist approach to Grounded Theory, including the leading principles of saturation and coding (Charmaz, 2010). I achieved deep familiarity with the researched phenomena but did not superimpose theories on the data. In Charmaz's (2010) words, "preconceived theoretical concepts may provide starting points for looking at the data but they do not offer automatic codes for analyzing these data" (p. 68). I coded the three data sources through two cycles and I ran descriptive statistics via NVIVO software. Through the first coding cycle all patterns that emerged resonated with the Universal Design for Learning (UDL) model, which is relevant to our institutional commitment to diversity, indigeneity, and serving our variegated student population (Rao, in press).

[The] methodology follows a constructivist approach to Grounded Theory, including the leading principles of saturation and coding

Therefore, I decided to conduct the second coding cycle according to UDL checkpoints. I indicated percentages of how many times each of the nine UDL guidelines emerged from the data; at the same time, I was mindful that all strategies are equally important and used in different contexts (Rao, in press). My student assistant also coded the data, providing a learner's perspective in the study (Desiree, Hernández, & Berumen, 2018; Driscoll & Wood, 2007; O'Neill & Maguire, 2017). As we moved from course-level assessment reports to index cards and faculty's conversation, we reached a saturation point—no more new information emerged.

Theorizing Assessment in Relation to Quality and Equity Assurance in the Fourth Industrial Revolution

Quality assurance refers to embedding quality in all aspects of higher education, from student learning and development to administrative processes, rather than inspecting quality in unmonitored systems that have already been finished (Ryan, 2015). Quality within the higher education environment has historically been one of the most difficult specifications to measure (Neubauer, 2019). Globalization challenges all facets of higher education, including its efforts to develop systems of quality assurance that are useful and sustainable in the 21st century (Neubauer, 2010). Over the past two decades, reductionist approaches have established quantitative tactics, such as the ranking phenomenon (Neubauer, 2018). Yet, what Klaus Schwab defines as the Fourth Industrial Revolution or Work 4.0 is already pushing quality in higher education towards

more complex progressions (cited in Neubauer, 2018). Implications will likely affect the manner in which university teaching is organized, due to student mobility, interdisciplinary demands, and need for self-learning skills as well as learning that is immersive, interactive, and responsive to students (Neubauer, 2019).

Student variability is possibly the most complex feature in the 21st century classroom. Globalization is not only an "intensifier of interdependence, it is a multiplier and magnifier of differences" (Hershock, 2010, p. 30). Each individual identity holds both cosmopolitan (i.e., humanist and global) and grounded (i.e., local and national) affiliation (Kahn, 2004; Logli, 2016; Nilan & Feixa, 2006). Within education, variety tends to relate to gender, ethnicity, religion, and social class as well as learning styles (Hershock, 2010; Hershock, Mason, & Hawkins, 2007). Concerns of access, survival, output, and outcome that relate to differentiated student populations are widespread (Farrell, 2007; Hawkins, 2011). Underrepresented students' probability of getting into college, completing their degree, learning the same knowledge and living relatively similarly post-graduation lives is slimmer compared to students who more traditionally attend higher education (Farrell, 2007).

Student variability is possibly the most complex feature in the 21st century classroom

Hershock (2010, 2012) expands the lens of analysis on diversity. Variety is "a quantitative index of simple multiplicity that connotes things simply being-different" (Hershock, 2010, p. 35). University campuses can host varied student populations, program offerings, and partnership types while remaining mere coexistence. Like zoos, their variety is externally imposed. By contrast, diversity is "a qualitative index of self-sustaining and difference-enriching patterns of mutual contribution to shared welfare" (Hershock, 2010, p. 35). Diversity refers to "the extent to which differences are activated as the basis of meaningful contribution to sustainably shared flourishing" (Hershock, 2012, p. 44). Diversity requires making differences as the basis for mutual contribution. Educating for diversity is focused on "discerning how most effectively and sustainably to enable the differences of each to make a difference for all" (Hershock, 2010, p. 38). It means "shifting the locus of concern from how much we differ-from each other to how we might best differ-for one another" (Hershock, 2010, p. 38). Like ecosystems, their diversity can only rise from within.

This theory of diversity can find two entryways into the assessment field—through Assessment 2.0 because it is contextual, flexible, and open (Metzler & Kurz, 2018) as well as culturally responsive assessment because it zooms into matters of equity (Montenegro & Jankowski, 2017b). Assessment 2.0 is designed to "supplement the assessment work already being done" and to be "organic"—growing naturally from faculty and staff's professional judgment and experience, rather than over-imposed structures and linear procedures commonly followed in standard assessment practice (Metzler & Kurz, 2018, p. 4). Its premise is that assessment must lead to action—assessment should not be done unless there is real possibility and openness to instructional and institutional change in the students' best interest. Campuses are filled with data, but data without rigorous analysis and usage are useless (Allen, 2004; Kuh et al., 2014).

Assessment 2.0 re-envisions assessment by using awareness of George Ritzer's theory of rationalization in order to avoid its most problematic expressions (Metzler & Kurz, 2018, p. 5). As all other system of rationalization, assessment has four interrelated dimensions. First, efficiency refers to streamlining processes to fulfill the end goal; yet, the risk is to diminish learning to shallow objectives. Second, calculability emphasizes measurable data; yet, the danger is that quantity takes precedence over quality. Third, predictability is concerned with limiting variability; yet it can brush off fine and contextual distinctions around student learning. Fourth, control aims to manage the volatility that is part of human dynamics; yet it can veer to collection of meaningless data. Without rationalized processes, chaos would prevail. However, the more a structure is rationalized, the greater the risk that its rationalization will undercut its own purpose, leading to lack of validity, reliability, reflection, and usage in assessment data. The Assessment 2.0 paradigm avoids the irrationality of rationality by infusing the four dimensions of rationalization with bottom-up processes, collective meaning-making within departments, and organic opportunities to provide assessment data, whether via formative, summative, quantitative, or qualitative approaches.

Culturally responsive assessment reveals the flawed assumption that "while there are multiple ways for students to learn, students need to demonstrate learning in specific ways for it to count" (Montenegro & Jankowski, 2017b, p. 6). Where one assessment approach is dominant there is a risk that it will not accommodate individual learning preferences (O'Neill & Maguire, 2017). Assessment approaches have a powerful impact on students' behavior, engagement, and attendance as well as institutional performance. For equity gaps to be addressed, the entire institution needs to explore resolutions to support student success—from pedagogies to assessments that foster inclusion while upholding high standards (Montenegro & Jankowski, 2017b). Equity-based approaches do not benefit underrepresented students only, but all students (Finley & McNair, 2013). For example, a "culturally responsive assessment" involves students throughout the entire assessment process, develops evaluation tools that are appropriate for different learners, uses results to improve the academic experience of all students, and disaggregates the data to understand the student population (Montenegro & Jankowski, 2017b). Once we know who our students are we can tailor assessment processes to better encourage their flourishing.

Institutions have started using a variety of approaches to get a more holistic picture of student learning and development rather than relying solely on exam-taking abilities. In the United States, community colleges and Minority-Serving Institutions integrate mainstream measures with locally developed instruments that are better suited to gauge their learning (Montenegro & Jankowski, 2015). Various campuses have centered assessment work on social justice principles through democratic, participatory, inclusive, affirming, and collaborative practices (Desiree et al., 2018). Examples from other English-speaking countries include (a) the University of East London, which gives the option to do a presentation, poster, or debate; (b) the University of Dublin, where students are able to make a poster instead of taking an exam; and (c) a study in Canada that has shown that students who took advantage of the option to add a term project through preparing a mini-class or participating in community service improved their performance (Montenegro & Jankowski, 2017b). Allowing students to choose how they are evaluated improves student engagement, achievement, and the quality of the learning experience while addressing student variability (Gosselin & Gagné, 2014).

The Universal Design for Learning (UDL) approach provides a useful application of culturally responsive assessment. UDL is based on the premise that variability among learners is the norm and individuals can become expert learners in varied ways—there is no one path to mastery (Rose & Gravel, 2009). To design for variability instructors can begin by identifying common barriers to learning, students' preferences, and specific needs for supports (Hehir, 2009). By being mindful of these factors, teachers can design instruction and assessment from the outset for a broader range of learners (Rao, in press). The UDL model is comprised of three main principles—representation, action and expression, and engagement. Each principle has three guidelines (nine in total) and each guideline has a series of checkpoints (31 in total) that proactively build in flexibility, choice, and scaffolds as well as other pedagogical practices to facilitate the learning experience for all.

Institutions have started using a variety of approaches to get a more holistic picture of student learning and development rather than relying solely on exam-taking abilities

Many assessment scholars suggest similar practices, but the UDL guidelines explicitly point out their relevance for student variability and provide a comprehensive model. For example, the literature on assessment and equity highlights the importance of fostering students' self-esteem, motivation, engagement, critical thinking, and leadership skills by integrating their indigenous cultural practices, appropriate language, and unique learning interests (Montenegro & Jankowski, 2015). Research on assessment and learning has also demonstrated the high impact of practices such as diversity and global learning, collaborative assignments, research opportunities, and all forms of service-learning (Kuh, O'Donnell, & Reed, 2013). Similarly, students indicated group work, application of knowledge, interaction with peers, and real-life connection as the activities that engaged them the most (Ewell, 2009). Studies on human learning suggest the importance of practice at retrieval, varying conditions under which learning takes place, and re-representing information in alternative formats as well as integrating theory with practice (Halpern & Hakel, 2003). Scaffolding should infuse all aspects of teaching from

instruction to assessment, for instance through sequenced lesson plans, rubrics that illustrate criteria, opportunities for student self-reflection on learning, and assignments that build on one another and prepare the student for a culminating demonstration (Hutchings et al., 2014).

Culturally Responsive Assessment 2.0 in Practice: Three Approaches

Three sets of data are the focus of this paper: (a) course-level assessment reports, (b) a campus-wide activity on student engagement, and (c) discussions around assessment, including gains and gaps in student learning as well as prospective improvements in all facets of institutional performance. The results consistently confirmed each other across the three datasets; therefore, I will present the two key findings from the three datasets collectively.

Diversified Assessment Methods

All three data sources revealed that faculty members diversify their assessment methods within a course but students are rarely given a choice on how to be assessed

All three data sources revealed that faculty members diversify their assessment methods within a course but students are rarely given a choice on how to be assessed. According to the assessment reports, 29% of the CLOs were assessed through exams, 21% through embedded questions, 18% through lab tasks, 9% through activities, 8% through presentations, 7% through projects, 4% through papers, and 3% through practica. During the discussions faculty shared specific examples of current practices and upcoming plans around culture-based assessment in their classroom, based on what they learned about different assessment methods. They expressed interest in diversifying their assessment methods further and providing choice to students on how to be assessed, as they realize that our diverse population benefits from it. Deeper considerations about student learning and matters of equity lie underneath this combination of assessment methods. Two examples clarify this multifaceted approach.

On the first day of class, a faculty member asks students to fill out the "About You Questionnaire," which allows her to get to know her students and to draft assessment methods accordingly (Hartline, 2018a). For instance, (a) if some students do not have a computer she allows hand-written submissions; (b) if some students do not have a printer, she allows digital submissions; (c) if some students have dyslexia, she allows submissions via audio recording; (d) if some students deal with anxiety disorders, she replaces whole-class presentations with group work; (e) if some students are veterans with medical and readjusting challenges, she views behavior that may seem otherwise antisocial (putting their head in their hands or standing up at odd times) as a response to easily triggered migraines and physical pain; (f) if some students are not native English speakers, she slows down and interprets roadblocks from a cultural lens; (g) if students take the bus, she ensures to open her classroom door ahead of time, so that they're not waiting in the hallway, and to end class on time, so they do not miss their public transportation (Hartline, 2018b).

Another department conducted a longitudinal study on its assessment methods and found that they are all informative but in different ways (Patterson, 2018). First, students' self-reported Knowledge Surveys provide a good overview of what they are learning, primarily in terms of content areas, because in the cultural context of Hawaii students appear to be modest in their self-reporting. Second, embedded assessment adds more specific understanding about students' analytical skills. Third, faculty's reflections on outcomes solidify the identification of better teaching strategies.

These findings are in line with principles of culturally responsive assessment (Gosselin & Gagné, 2014; Montenegro & Jankowski, 2015; O'Neill & Maguire, 2017). Faculty members diversify their assessment tools to accommodate for student variability. They also contextualize their evaluations within the reality of Hawaii and integrate mainstream measures with locally developed instruments that are better suited to gauge learning. The faculty's autonomy in designing assessment methods also resonates with the Assessment 2.0 model (Metzler & Kurz, 2018). Faculty members prefer embedded assessments to use data sources that already exist and supplement the assignments already being done. They freely choose either formative or summative approaches, by drawing from their professional judgment and experience, rather than over-imposed standardized assessment plans. They find faculty development, fresh resources,

and collegial conversations around assessment useful as they strategize on how to improve student learning.

Teaching for Student Variability

All UDL principles emerge as relevant and, as I discuss below, three guidelines are especially intriguing. Faculty use a variety of inclusive practices to enhance student learning, by connecting academic and student affairs (UDL guideline 6), providing options for student persistence (UDL guideline 8), and optimizing students' motivation and coping skills, mainly through relation-building (UDL guideline 9). Interestingly, over the past years the campus has been increasingly concerned about student retention and providing student support has been a priority, mainly through a caring environment and collaboration between instructional and noninstructional units.

Table 1
Synopsis of findings within UDL model

	I. REPRESENTATION (guidelines 1-3)	II. ACTION/ EXPRESSION (guidelines 4-6)	III. ENGAGEMENT (guidelines 7-9)
Course-level: Assessment reports	**47%** **(312 of 654 CLOs)** •UDL guideline 2: 14% •UDL guideline 3: 33%	**30%** **(193 of 654 CLOs)** •UDL guideline 4: 4% •UDL guideline 5: 6% •UDL guideline 6: 20%	**23%** **(149 of 654 CLOs)** •UDL guideline 7: 1% •UDL guideline 8: 20% •UDL guideline 9: 2%
Institutional level: Student engagement practices	**4%** **(10 of 233 practices)** •UDL guideline 2: 1% •UDL guideline 3: 3%	**29%** **(68 of 233 practices)** •UDL guideline 4: 6% •UDL guideline 5: 7% •UDL guideline 6: 16%	**67%** **(155 of 233 practices)** •UDL guideline 7: 5% •UDL guideline 8: 21% •UDL guideline 9: 40%

First, providing options for executive functions (UDL guideline 6) is at the forefront of faculty's considerations—they mention it 20% of the time in their course-level assessment reports, 16% of the time in the campus-wide index cards activity, and in discussions. They support planning and strategy development by connecting students to campus resources "that can help overcoming their challenge" (e.g., writing center, testing center, Hawaiian center, veterans center, student success center, library, academic counseling, career services, tutoring, transition coordinator, peer mentors, noncredit ESL classes, financial aid). Bridging classrooms with services has been a focus on campus over the past years and faculty have been responsive by inviting guest speakers from student affairs in their courses, sending students on a scavenger hunt to key spaces on the first day of class, and referring students to a variety of supports. In addition, they guide appropriate goal setting by organizing the syllabus thoroughly, sending various types of e-newsletters to students before the start of the semester, and adjusting the course pace—they either "set quick turnaround time" or "allow more time" depending on the circumstances. They also facilitate managing information by "putting great thought" into structuring mind mapping, practice sheets, and transition projects. Furthermore, they enhance the capacity for monitoring progress by "having an assignment where students plot a course outline to reach their end goal," "keeping students accurately updated," and "correcting each deficiency before moving on to the next project" through outside-of-class optional review sessions, in-class practice exams with samples, and graded pre-quizzes.

Second, providing options for persistence (UDL guideline 8) is another faculty priority— they mention it 20% of the time in their course-level assessment reports, 21% of the time in the campus-wide index cards activity, and in discussions. They heighten the salience of goals and objectives by engaging students in activities that are relevant for their lives (e.g., field studies,

They heighten the salience of goals and objectives by engaging students in activities that are relevant for their lives

outside-of-the-classroom projects, guest speakers, analyses of real-world problems and current events). In addition, they vary demands and resources to optimize challenge—"I diversify my teaching strategies, I switch mode every 10 minutes to support each learning preference." They also foster collaboration by engaging families when appropriate and integrating group activities (e.g., ice breakers, get-togethers, partnering in problem solving) so students "get to know one another—who they are and what their interests are—and make discoveries, so they are happy to return to class because their friends are there." Faculty encourage student participation in campus life (e.g., student clubs, social projects, leadership opportunities) so "they experience values like community and compassion, and can be the positive change that ripples around." Moreover, they increase mastery-oriented feedback by using comments like "the essay would be better with punctuation," rather than "you need to work on punctuation," through lab follow-up, discussions on assignments, well-defined rubrics, peer mentorships, and learning communities where students "share their mistakes, discoveries, and learn from each other."

Third, providing options for self-regulation (UDL guideline 9) was the highlight of the index cards activity—faculty and staff mentioned it 40% of the time. One reason for this predominance is that the ninth UDL guideline falls under the UDL engagement principle and the activity was about purposeful engagement. Yet, discussions confirmed its importance. Faculty members promote expectations and beliefs that optimize motivation, by assisting them to solve issues instead of passing them along, taking the time to understand their needs, and "teaching them how to be students, which can translate to all courses" (e.g., how to take notes, quizzes). In addition, they facilitate personal coping skills and strategies by providing personal stories and professional mentorship around their passions, dreams, difficulties, and strengths. For instance, they make an effort to learn students' stories (e.g., show up early to class, create talk story/snack time, conduct "I wish my teacher knew" exercise) and provide professional guidance (e.g., provide letters of recommendation, share networking, revise job applications, organize mock interviews, encourage students to think about "short- and long-range goals within the industry"). They also develop self-assessment and reflection through one-minute surveys (e.g., What did you learn today? What did you have more questions about?), end-of-the-semester meetings, course evaluations, and involving them in rubric development.

At this point, I want to touch upon all other culturally responsive strategies that emerged from the data. Their numerical weight is more limited—yet still substantial, considering that percentages are spread across the nine UDL guidelines—and their qualities remain impactful across all three datasets. Faculty clarify vocabulary and promote understanding across languages (UDL guideline 2, mentioned 14% of the time in course-level assessment reports). They provide options for comprehension (UDL guideline 3) by explaining "stories behind place names," "replacing textbook examples with local examples," and using familiar images like a rainbow to capture student attention around salient scientific characteristics (mentioned 33% of the time in course-level assessment reports). They "help students make connections with class content"— through hands-on activities in order "to show students that what they are learning is practical, important, and related to both local and global perspectives" (e.g., Malama Aina or Take Care of the Land days).

Faculty optimize access to technologies (UDL guideline 4), for example by "integrating online research skills into some of the hands-on assignments" and "giving extra credit for early online communication" (mentioned 6% of the time in the campus-wide index cards activity). They also provide options for expressions (UDL guideline 5), such as through in-depth engaging questions, experimental learning, and industry networking in order "to create space for students to became part of the program in significant ways" (mentioned 7% of the time in the campus-wide index cards activity). Providing scaffoldings and opportunities for practice is key, not just for content absorption but also for learning the skill "on how to learn more."

Faculty optimize autonomy (UDL guideline 7) by allowing students to select topics, learning tools, and assignments (mentioned 5% of the time in the campus-wide index cards activity). For example, students chose "the ideas and activities they wanted to plan and share with

classmates," create ground rules for class, and are in charge of opening and closing the lab. They minimize distractions, for instance, by creating a safe space, respectful learning atmosphere, and friendly environment "so that students feel comfortable to interact with each other and help each other learn better."

These findings confirm some of the guidelines of culturally responsive assessment (Finley & McNair, 2013; Montenegro & Jankowski, 2017b). Faculty members involve students throughout the assessment process, draw from culture-based lenses, and use assessment results to improve learning of all students. These results also follow the Assessment 2.0 model as our institution has started using a variety of approaches to get a more holistic picture of student learning, including assessment reports that make space for qualitative evidence, including thick and rich reflections (Metzler & Kurz, 2018).

Culturally Responsive Assessment 2.0: Implication for Future Directions

Today's educational challenges are not problems to be solved but rather predicaments to be resolved (Hershock, 2010, p. 31). Problems are associated with failures of existing practices and are solved by developing improved means, which lead to ends that we intend to continue pursuing. Predicaments arise when changing circumstances bring the awareness of conflicts among our own values and are resolved by creating new meaning, which activates shared commitments. In considering the direction of assessment paradigm changes, matters of difference and equity should not be seen as side effects of education but rather values that can be effectively promoted only when infused into the full spectrum of educational activity.

Similar to the case studies featured in Baker, Jankowski, Provezis, and Kinzie (2012), this pilot project sees room for development and does not conclude that it is "yet arrived" (p. 6). Assessment is never completed but it continuously advances through an ongoing process in which closing the loop opens new inquiries. Some of the limitations of this study include its pilot nature, case-study scale, and focus on faculty's direct assessments rather than also including students' perspectives. From a theoretical stance, I plan to contribute to studies on assessment approaches first in relation to indigenous contexts and later beyond the dominant Anglo-American axis. From a practical stance, my institution plans to continue implementing assessment approaches that are flexible (as suggested by Metzler & Kurz, 2018) as well as culturally responsive (as suggested by Montenegro & Jankowski, 2017b). Rationalized and yet flexible processes will keep leading assessment efforts.

First, with regard to efficiency, a new curriculum process will allow a more flexible procedure for updating and linking outcomes, while also maintaining the principles that outcomes should be concise and clear for all students. A new Assessment Management System (AMS) will allow data disaggregation by student background and give faculty options, by asking some brief standardized questions and creating a space for additional data on specific interests (Baker et al., 2012). Specific interests will also evolve, for example, around place-based approaches, which remain the essence of our campus as demonstrated by our findings (Baker et al., 2012; Montenegro & Jankowski, 2017b).

With regard to efficiency, a new curriculum process will allow a more flexible procedure for updating and linking outcomes, while also maintaining the principles that outcomes should be concise and clear for all students

Second, with regard to calculability, the new AMS will allow desegregated analytics to investigate how the assessment might affect all students, benefit certain populations, and hinder others. Qualitative methodologies will continue to illuminate the complex predicaments and resolutions that lie underneath numerical values. Student voices will also become part of the data triangulation, for instance, by integrating Community College Survey of Student Engagement (CCSSE) results as well as many other student surveys that disseminate on campus (Driscoll & Wood, 2007).

Third, predictability will appear through embedding assessment into institutional processes such as program reviews, decision-making by campus-wide committee, and external documents so that assessment efforts are optimized (Baker et al., 2012). For example, both applications and reports for Title III federal grants in support of Hawaiian values need culturally

responsive assessment data. Information from assessment results, including success stories and action plans, should be distributed more widely to both internal and external audiences (Baker et al., 2012). For example, a top administrator integrated the analysis of the index cards into her presentation at the planning council. Predictability will also motivate the drafting of new assessment policy and procedures in order to bring together various perspectives and communicate clearly leading principles—above all assessment data needs to be collected, analyzed, discussed, and used toward improvement in student learning and institutional performance, regardless of specific details.

Fourth, control can frame more widely the spaces where dialogues about student learning and institutional performance become usable data. Conversations about improving student success take place constantly in numerous informal channels and those voices should be included in datasets (Baker et al., 2012). A committee has just stepped forward to become the body to examine five-year program reviews, encourage the integration of program-level assessment analyses, and to find ways to support the department. Fresh efforts have started to include nontenure-line faculty in campus initiatives such as curriculum mapping, optional orientation, and faculty training, including the assessment series (Kezar & Maxey, 2014). The monthly assessment series tries to meet various accessibility preferences—in addition to the face-to-face sessions, I post on the assessment page the full PowerPoint, a one-page handout, and a five-minute online tutorial with captures.

The study finds that assessment data emerge spontaneously across numerous spaces on campus, can be collected in flexible manners, and can be analyzed through an equity lens in order to support a diverse student population.

In conclusion, offering an assessment framework toward a more equitable higher education landscape motivates this study. This paper argues for organic initiatives that supplement the assessment work already being done, by blooming naturally from faculty and staff's expertise while also pollinating each dimension of the assessment with culturally responsive attention (Metzler & Kurz, 2018; Montenegro & Jankowski, 2017b). The study finds that assessment data emerge spontaneously across numerous spaces on campus, can be collected in flexible manners, and can be analyzed through an equity lens in order to support a diverse student population. In particular, the results point out that faculty members are implementing inclusive practices in their instruction and assessment, and closely collaborate with units in student affairs to engage all learners. This flexibility is endorsed by accrediting agencies, which do not prescribe narrow templates but rather leave institutions free to sculpt their own assessment image (Cain, 2014). Accreditation protects institutional autonomy, academic freedom, and institutional diversity (American Council on Education, 2012). Assessment rightly conducted "asks faculty to work together as colleagues to assess student work fairly by criteria respected in the field and to share their knowledge of student strengths and weaknesses, in order to improve curriculum, pedagogy, and other factors affecting learning" (Cain, 2014, p. 12).

References

Accreditation Taskforce. (2018). Honolulu Community College: Institutional self-evaluation report. Honolulu, HI: Honolulu Community College.

Allen, M. (2004). Assessing academic programs in higher education. Bolton, MA: Anker Publishing Company.

American Council on Education. (2012). Assuring academic quality in the 21st century: Self-regulation in a new era. Washington, DC: American Council on Education.

Arbuckle, J. (2018). Honolulu Community College: Fact book. Honolulu, HI: Honolulu Community College.

Baker, G. R., Jankowski, N. A., Provezis, S., & Kinzie, J. (2012). Using assessment results: Promising practices of institutions that do it well. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Barr, R., & Tagg, J. (1995). From teaching to learning—A new paradigm for undergraduate education. Change: The Magazine of Higher Learning, 27(6), 12–26.

Cain, T. (2014). Assessment and academic freedom: In concert, not conflict. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Charmaz, K. (2010). Constructing grounded theory: A practical guide through qualitative analysis. Thousand Oaks, CA: Sage.

Desiree, Z., Hernández, I., & Berumen, J. (2018). Assessment and social justice: Pushing through paradox. New Directions for Institutional Research (177), 1–144.

Driscoll, A., & Wood, S. (2007). Developing outcomes-based assessment for learner-centered education: A faculty introduction. Sterling, VA: Stylus.

Ewell, P. (2009). Assessment, accountability and improvement: Revisiting the tension. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Farrell, J. (2007). Equality of education: A half-century of comparative evidence seen from a new millennium. In R. Arnove & C. A. Torres (Eds.), Comparative education: The dialectic of the global and the local (pp. 129–150). Lanham, MN: Rowman & Littlefield.

Finley, A., & McNair, T. (2013). Assessing underserved students' engagement in high-impact practices. Washington, DC: Association of American Colleges and Universities.

Geertz, C. (1973). The interpretation of cultures: Selected essays. New York, NY: Basic Books.

Gosselin, J., & Gagné, A. (2014). Differentiated evaluation: An inclusive evaluation strategy aimed at promoting student engagement and student learning in undergraduate classrooms. Toronto, Canada: Higher Education Quality Council of Ontario.

Halpern, D., & Hakel, M. (2003). Applying the science of learning to the university and beyond: Teaching for long-term retention and transfer. Change: The Magazine of Higher Learning, 35(4), 36–41.

Hartline, E. (2018a). About You Questionnaire. Honolulu Community College. Honolulu, HI.

Hartline, E. (2018b). Creating an assessment plan: Understanding students as the foundation of assessment [PowerPoint slides]. Honolulu, HI: Honolulu Community College.

Hawkins, J. (2011). Variations on equity and access in higher education in Asia. In D. Neubauer & Y. Tanaka (Eds.), Access, equity, and capacity in Asia-Pacific higher education (pp. 22–44). New York, NY: Palgrave Macmillan.

Hehir, T. (2009). Policy foundations of universal design for learning. In D. Gordon, J. Gravel, & L. Schifter (Eds.), A policy reader in Universal Design for Learning (pp. 35–45). Cambridge, MA: Harvard Education Press.

Hershock, P. (2010). Higher education, globalization and the critical emergence of diversity. Paideusis, 19(1), 29–42.

Hershock, P. (2012). Mobility, markets and equity in higher education: Match or mismatch? In D. Neubauer & K. Kuroda (Eds.), Mobility and migration in Asian Pacific higher education (pp. 21–37). New York, NY: Palgrave Macmillan.

Hershock, P., Mason, M., & Hawkins, J. (2007). Changing education: Leadership, innovation and development in a globalizing Asia Pacific. Hong Kong: Springer.

Hutchings, P., Jankowski, N., & Ewell, P. (2014). Catalyzing assignment design activity on your campus: Lessons from NILOA's assignment library initiative. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Kahn, J. (2004). Introduction: Identities, nations and cosmopolitan practice. Paper presented at the Identities, Nations and Cosmopolitan Practice: Interrogating the Work of Pnina and Richard Werbner, Singapore.

Kezar, A., & Maxey, D. (2014). Student outcomes assessment among the new non-tenure-track faculty majority. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Kuh, G., Ikenberry, S., Jankowski, N., Cain, T., Hutchings, P., & Kinzie, J. (2014). Using evidence of student learning to improve higher education. San Francisco, CA: John Wiley & Sons.

Kuh, G., O'Donnell, K., & Reed, S. (2013). Ensuring quality and taking high-impact practices to scale. Washington, DC: Association of American Colleges and Universities.

Logli, C. (2016). Cultural studies of education: Filming fluid subjectivities in Indonesian universities. Pedagogy, Culture & Society, 24(2), 177–190.

Metzler, E. T., & Kurz, L. (2018). Assessment 2.0: An organic supplement to standard assessment procedure. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Montenegro, E., & Jankowski, N. (2015). Focused on what matters: Assessment of student learning outcomes at minority-serving institutions. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Montenegro, E., & Jankowski, N. (2017a). Bringing equity into the heart of assessment. Assessment Update, 29(6), 10–11.

Montenegro, E., & Jankowski, N. (2017b). Equity and assessment: Moving towards culturally responsive assessment. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Neubauer, D. (2010). Ten globalization challenges to higher education quality and quality assurance. Evaluation in Higher Education, 4(1), 13–37.

Neubauer, D. (2018, September 24). Reassessing the nature and dynamics of student mobility within Asia Pacific higher education. Paper presented to the UMAP International Forum, Osaka, Japan.

Neubauer, D. (2019, September 8-10). Recalculating higher education in the Asia Pacific region within the emerging Fourth Industrial Revolution. Paper presented to the CHER Conference: Reassessing Higher Education in the Asia Region, Lingnan University, Hong Kong.

Nilan, P., & Feixa, C. (2006). Global youth? Hybrid identities, plural worlds. London, UK: Routledge.

O'Neill, G., & Maguire, T. (2017). Developing assessment and feedback approaches to empower and engage students: A sectoral approach in Ireland. Dublin, Ireland: National Forum for the Enhancement of Teaching and Learning in Higher Education.

Patterson, P. (2018). Utilizing assessment data: A two-track approach [PowerPoint slides]. Honolulu, HI: Honolulu Community College.

Rao, K. (2019). Instructional design with UDL: Addressing learner variability in college courses. In S. Bracken & K. Novak (Eds.), Transforming higher education through Universal Design for Learning: An international perspective (pp. 115-130). New York, NY: Routledge.

Rose, D., & Gravel, J. (2009). Getting from here to there: UDL, global positioning systems, and lessons for improving education. In D. Gordon, J. Gravel, & L. Schifter (Eds.), A policy reader in Universal Design for Learning (pp. 5–18). Cambridge, MA: Harvard Education Press.

Ryan, T. (2015). Quality assurance in higher education: A review of literature. Higher Learning Research Communications, 5(4).

Abstract

Throughout the literature, there is ongoing discussion about what constitutes student engagement and how it is defined. Most metrics evaluate student engagement through behavioral engagement, rather than including other aspects such as cognitive or emotional engagement (Lester, 2013). Engaging in activities considered to be "high-impact practices" during undergraduate studies is related to academic success (e.g., Kuh, Cruce, Shoup, Kinzie, & Gonyea, 2007; Kuh, Kinzie, Schuh, & Whitt, 2005). The purpose of this pilot study was to examine the relationship between different types of student engagement (i.e., behavioral, cognitive, and emotional) and value-added scores in undergraduate students. Types of engagement (i.e., cognitive and psychological/emotional engagement) were observed through participation in student activities and programs, such as the Leadership Programs (LP) and Greek Life. Scores on the Collegiate Learning Assessment Plus (CLA+) were used to assess value-added, longitudinally comparing the change from students' first-year and senior-year administrations of the test. We sought to better understand the relationship between different types of student engagement and explore which of those may have influenced the value-added scores in undergraduate students.

AUTHORS

Jason C. Lyons and
Gina B. Polychronopoulos

Christopher Newport University

Exploring the Relationship between Types of Engagement and Value-added Scores in Undergraduate Students

In recent years, tuition in higher education has increased substantially (The College Board, 2019; U.S. Department of Education, National Center for Education Statistics, 2019), conceivably leaving prospective students to wonder what they will gain from pursuing a college degree. They may question whether what they learn in classes will help them in the job market (particularly at a liberal arts institution), and administrators and faculty members are tasked with helping students to understand the value of the undergraduate experience. Completing academic requirements and mastering course content are an important part of the equation; however, what students do outside the classroom (e.g., cocurricular activities, community engagement, and social clubs/organizations) may also factor into the value of one's undergraduate experience. Indeed, engaging in certain types of activities considered to be "high-impact practices" during undergraduate studies has been shown to be related to academic success (Kuh, Cruce, Shoup, Kinzie, & Gonyea, 2007; Kuh, Kinzie, Schuh, & Whitt, 2005). Researchers have long been exploring this phenomenon in an effort to answer the overarching question, "How do the activities that students engage in outside of the classroom influence their academic success, and to what extent?" To help answer this question, it is important to understand what student engagement means and how it has been studied.

Student Engagement

The role of student engagement in academic achievement has been studied for decades in the United States and began picking up momentum in the 1980s, with Alexander Astin's research on the relationship between student engagement and the educational experience. Around that time educational researchers began identifying "good" educational practices and investigating how they might be related to students' college experience (Axelson & Flick, 2011). Thereafter, the discussion progressed toward the challenge of defining the various dimensions of engagement, all of which are in continuous need of refinement, as well as how to measure the dimensions of student engagement.

CORRESPONDENCE

E-mail: jason.lyons@cnu.edu

One example of a measure for student engagement is the National Survey of Student Engagement (NSSE), which solicits self-report information (perceptions) from students about how often they engage in behaviors related to the student experience. Some examples include (but are not limited to) how often they participated in activities, how much time they spend studying or preparing for a course, and their frequency of using various resources on campus. In general, the NSSE focuses on a student's level of involvement as measured by frequency of behavior (i.e., behavioral engagement) and may not account for the emotional or cognitive/psychological aspects of engagement; therefore, it may be helpful to further refine the definition of engagement to better understand what aspects of student involvement may be influencing the student experience (Axelson & Flick, 2011).

Student engagement is a multidimensional construct, and as such, it is challenging to reduce it to singular dimensions and isolate its effects; they are often interrelated. According to Fredricks, Blumenfeld, and Paris (2004), most conceptualizations of "school engagement" included some combination of behavioral, emotional, or cognitive aspects. Using the Student Course Engagement Questionnaire (SCEQ), Handelsman, Briggs, Sullivan, and Towler (2005) revealed four distinct and reliable dimensions of student engagement through exploratory factor analysis: skills, participation and/or interaction, emotional, and performance (Handelsman, Briggs, Sullivan, & Towler, 2005). Behavioral engagement, therefore, may refer to actions such as participating in school-related activities, studying or completing assignments, adhering to rules of conduct, and involvement in athletics. Emotional engagement typically includes students' attitudes and feelings in relation to the institution, faculty, and other students, and feeling like one belongs in the institutional environment. Cognitive engagement (including psychological engagement) includes student motivation to learn and personal investment in learning. Fredricks et al. (2004) concluded that overall, the construct of engagement is multi-dimensional and warrants further exploration (Fredricks, Blumenfeld, & Paris, 2004). Lester (2013) further emphasized the importance of clarifying and refining the definition of student engagement, in order to better understand its influence on student outcomes (Lester, 2013).

As far as we can tell, defining engagement in terms of participation in programs that are specific to the university is something that has not been attempted. Therefore, in the interest of further exploring this construct, we observed cognitive, behavioral, and emotional engagement as defined by a student's participation in one of three groups: Honors Program (cognitive), Leadership Program (behavioral), and Greek Life (emotional). These engagement groups are specific to the university, although there may be similar programs at other universities. The following sections describe these student activity groups and how each may reflect a different type of student engagement, as defined in this pilot study.

Honors Program: Cognitive engagement. Students who enroll in the honors program must maintain a GPA of 3.5 or above throughout the course of their studies. In addition to completing the requirements of their major discipline areas, they also are expected to enroll in interdisciplinary seminars, engage in advanced inquiry courses, perform civic engagement activities, engage in research, study abroad, learn a second language, and complete additional requirements (e.g. honors portfolio). Student progress is monitored throughout their academic tenure. These students are part of a challenging, high-achievement program and are expected to maintain their cognitive engagement above and beyond the average student.

Leadership Program (LP): Behavioral engagement. The LP is a selective program in which students are expected to engage in several areas on campus and in the community. They must adhere to a higher standard of behavioral conduct, engage in development, leadership, success, and experiential opportunities, and maintain a minimum GPA of 2.75 for the lowest level of scholarship (3.3 for the highest level). At the end of their studies, LP students must submit a portfolio that summarizes and synthesizes the personal importance of engaging in the LP program activities throughout their undergraduate experience. These students are expected to maintain a high level of behavioral engagement throughout the course of their studies.

> **Student engagement is a multidimensional construct, and as such, it is challenging to reduce it to singular dimensions and isolate its effects**

Greek life: Emotional engagement. Membership in a nationally recognized fraternity or sorority (by the North American Interfraternity Conference, Interfraternity Council, National Pan-Hellenic Council, or National Panhellenic Conference) is referred to as Greek Life on campus. Greek Life activities are student-led with faculty oversight, and members are recruited each year. Student members are expected to maintain a minimum GPA of 2.06 and at least 12 credit hours. Greek Life students are actively involved on campus and in the community. Emotional engagement was observed as membership in Greek Life because students belonging to this group are interpersonally connected within the group through social events, residential living, and service activities.

Institutional Value-Add

Measuring the value of an undergraduate education can be a daunting challenge, as there are a wide range of aspects that can be considered valuable to a student in his or her educational experience and each of these aspects can be difficult to isolate. Aside from the knowledge and skills gained in the classroom, participating in activities outside the classroom may be just as important. Sometimes, measures for student engagement (such as the NSSE) are employed to broaden the picture of the student's experience and accompany measures of student achievement (e.g., graduation rates and retention, persistence, and GPA). Lodge and Bonsanquet (2014) discussed the difficulty in finding valid (and inexpensive) ways to determine the quality of a student's learning experience. They noted that institutions often use student opinion or satisfaction surveys to assess the perceived value of their education; however, the quality of learning is far more complex (Lodge & Bonsanquet, 2014).

Other institutions may measure learning outcomes achievement as part of their calculation of institutional value-add, i.e., what students gain as part of their experience at a university or higher education setting. Douglass, Thomson, and Zhao (2012) explored different methods of evaluating learning outcomes, including self-reported student gains and standardized testing. They discussed the wide use of the Collegiate Learning Assessment (CLA) in the United States as a measure of value-add. Indeed, the CLA offers this metric in the institutional reports they distribute; however, the authors argue that using the CLA as the only measure of value-add may not capture the full effect at large and complex universities, noting that it may provide more meaningful information at small, liberal arts institutions (Douglass, Thomson, & Zhao, 2012). In an effort to respond to the increasing need to measure student learning at the institutional level, many institutions (including the institution of the present study) have taken to administering the

Collegiate Learning Assessment Plus exam (CLA+). Collegiate Learning Assessment Plus (CLA+). The CLA+ is a nationally-normed, value-added approach to the measurement of higher-order thinking skills developed by the Council for Aid to Education (CAE) in 2002 (originally known as the Council for Financial Aid to Education, 2018). The CLA+ looks at scientific and quantitative reasoning, critical reading and evaluation, critiquing an argument, analysis and problem solving, writing effectiveness, and writing mechanics with the ultimate goal of providing a summative measure of the value added by the school's programs. The overarching goal is to offer information that will assist the institution with understanding how well their students are improving over time in comparison with students at other institutions (Klein, Benjamin, Shavelson, & Bolus, 2007).

The CLA+ allows institutions to benefit from a model of continuous improvement by serving as an additional visual tool within the education process. The value-added approach taken by the CLA+ allows institutions to meaningfully relate student growth at their institution to the growth of students across other institutions with a similar entering academic ability (EAA). This offers an equitable account of a school's contribution to learning and to the development of students' higher-order thinking skills.

It is important to note that the process CAE uses to calculate the value-added score has recently changed, posing potential challenges for some institutions. Prior to the 2016-2017 academic year, value-added scores for the CLA+ were calculated by obtaining the difference

Aside from the knowledge and skills gained in the classroom, participating in activities outside the classroom may be just as important

between an institution's Actual Senior Mean CLA+ Scores and Expected Senior Mean CLA+ scores. Expected scores were determined by the typical performance of seniors at institutions testing similar samples of students. These samples were identified based on senior EAA scores and mean first-year performance on CLA+. Finally, the EAA was determined based on one of three sets of scores: (a) combined SAT Math and Critical Reading, (b) ACT Composite, or (c) Scholastic Level Examination (SLE) scores, as stated in the institutional report drafted and distributed to our institution by CAE. After the 2016-2017 academic year, the methods used to calculate the Expected Senior Mean CLA+ Score changed: the Expected Senior Mean CLA+ Score was then determined by samples identified based on senior parental education scores and mean first-year student performance (of that same year) on the CLA+.

Each year, we administer the CLA+ to either first-year students in the fall semester or seniors in the spring semester. The spring 2018 administration of the CLA+ was the first longitudinal cohort, i.e., those who took the exam as first-year students in 2014 and were taking it for the second time. This provided us the opportunity to measure how each individual student improved in their academics and account for their level of engagement at the institution. For this particular cohort, the value-add score (as defined by CAE using the calculations noted above) was not available to our institution because we do not administer the CLA+ to both first-year students and seniors in the same year. Therefore, in the pilot study, we used the CLA+ Total Scores at First Year and Senior Year (longitudinal, matched cases) to explore the relationship between specific types of student engagement and institutional value-added scores.

We were interested in how participating in student activities may be related to the success of our students

Purpose of Study

We were interested in how participating in student activities may be related to the success of our students, in an effort to further support and promote those practices on campus. In particular, we wanted to explore whether participating in specific programs was related to value-add, as measured by the Collegiate Learning Assessment (CLA+) from the beginning of their undergraduate studies in First Year (Time 1) to their Senior Year (Time 2). The questions that we explored were as follows: Is there a difference in CLA+ scores for students who participate in a particular engagement group and those who do not? How do CLA+ scores vary over time for students who engage in these groups? The results of this preliminary study may offer potentially valuable insights to faculty and staff about how student participation in specific activity groups may have a potential role in their academic achievement and the value-add of attending the university.

Method

Participants

Undergraduate students attending a public liberal arts university were randomly selected to participate in this study and recruited by the Office of Assessment. A total of 200 students completed the CLA+ during their first year as students, and 119 of those students returned to complete the assessment for a second time four years later, in their senior year. After data cleaning for outliers and nonvalid assessments, the final number was 98 students who completed both test administrations. There were 29 male and 67 female students; two students declined to answer. Eighty-two students (84%) identified as "White, non-Hispanic," five (5%) as "African American/ Black," two (2%) as "Hispanic or Latino," two (2%) as "Asian," two (2%) as "Other," and five (5%) declined to answer. The sample was similar demographically to the overall first-year student cohort from which they were randomly selected.

A total of 17 students were identified as participating in the Honors Program (HP). Of those students, 14 were also in the LP and five also participated in Greek Life. Three students were HP only. A total of 33 students were members of the LP. Of those students, 14 were also HP, six were also members of Greek Life, and five were members of all three groups; 13 students were LP only. A total of 30 students were members of Greek Life. Of those students, 11 were also LP, 5 were members of all three groups, yielding 19 students who were Greek Life only. See Table 1 for a summary of how many students engaged in each group.

Table 1.
Number of students in each type of engagement activity group.

Engagement Activity	Honors	LP	Greek Life	No Group
Honors (n=17)	*3	14	5	--
LP (n=33)	--	*13	6	--
Greek Life (n=30)	--	--	*19	--
No Group	--	--	--	18

*Intersections of the same group indicate number of students involved in only that activity

There was considerable overlap in engagement activity participation. In particular, most of the Honors students in this sample were also participating in the LP and/or Greek Life groups. This yielded three students who were HP only, which was too small of a group to compare mean scores. Therefore, we did not examine this activity group (cognitive engagement) further in this pilot study.

Measures

Engagement. Behavioral engagement was observed as membership in the LP, as students are expected to participate in a minimum number of events and activities, as well as adhere to a higher standard of conduct throughout the course of their studies. Emotional engagement was observed as membership in Greek Life (i.e., fraternity or sorority), as students belonging to this group are connected within the group through social events, residential living, and service activities. Students were coded as LP/non-LP and Greek Life/non-Greek Life; they were coded as belonging to a group if they did not also belong to one of the other groups (i.e., LP only, Greek Life only) in an effort to keep the groups independent.

Value-add. The CLA+ specifically measures several learning domains: scientific and quantitative reasoning, critical reading and evaluation, critiquing an argument, analysis and problem solving, writing effectiveness, and writing mechanics. The CLA+ exam is divided into two sections: (a) Performance Task (PT) and (b) Selected-Response Questions (SRQ). The PT presents students with a real-world situation that requires them to address the issue, propose a solution to a problem, or recommend a course of action to resolve a conflict. The PT measures the following constructs, as defined by CAE: analysis and problem solving, writing effectiveness, and writing mechanics. Students are able to use a document library filled with a variety of reference materials to support their responses. The SRQ measures scientific and quantitative reasoning, critical reading and evaluation, and the ability to critique arguments by identifying logical flaws through a series of questions. Like the PT, students have a document library to draw information in support of their answers.

Procedure

The Office of Assessment administered the CLA+ to a random sample of incoming first-year students in the fall semester. Students were solicited via e-mail to register for one of the available time slots. All CLA+ exams were administered electronically in computer labs on campus, with faculty serving as proctors. Students had up to 60 minutes for the PT and 30 minutes for the SR tasks. Test results, including analyses of the PT score, the SRQ score and the total CLA+ score, were delivered to the university after the testing window. Those students were recruited to take the CLA+ for a second time during the spring semester of their senior year, prior to graduation. Student data regarding type of engagement activities was retrieved from the Offices of Institutional Research and Student Activities and matched accordingly.

Study Design

We began this study as a mixed factorial design, with the between-subjects factors being LP and Greek Life, and the within-subjects factors being time (CLA+ Time 1 and CLA+ Time 2); however, due to the small sample and uneven group sizes we have refrained from statements of significance. We have instead reported descriptive statistics of the means for each engagement

group to help shed light on which areas we may wish to further explore with a larger sample. We were interested in exploring the following research questions for this preliminary study:

1) To what extent did CLA+ scores vary over time in LP students and non-LP students? For students in Greek Life and non-Greek Life students?

2) How do CLA+ scores differ for students who engaged in different types of student activities (i.e., LP and Greek Life) as compared with students who did not engage in those activities?

It was anticipated that CLA+ scores would increase from Time 1 to Time 2, for students who participated in an engagement group. We also hypothesized that CLA+ scores at Time 2 would be significantly different for students who engaged in specific student activities (i.e., LP or Greek Life) and those who did not.

Data cleaning. Data cleaning involved two steps: (a) addressing extreme outliers for CLA+ scores (dependent variable), and (b) accounting for students who "clicked through" the test (e.g., finished the performance/writing task in less than 15 minutes and/or the selected response task in less than 10 minutes) on either test administration (first year or senior year). The resulting sample was n = 98 students who completed the CLA+ at both Time 1 and Time 2, with both distributions appearing normal in skewness and kurtosis. See Tables 2 and 3 for summaries of data distributional properties of CLA+ scores over time for each engagement activity group.

> We also hypothesized that CLA+ scores at Time 2 would be significantly different for students who engaged in specific student activities (i.e., LP or Greek Life) and those who did not

Table 2
Initial Data Distributional Properties for CLA+ Total Scores, LP and Non-LP

		LP Time 1	Non-LP Time 1	LP Time 2	Non-LP Time 2
N	Valid	13	85	13	85
	Missing	0	0	0	0
Mean		1176.54	1128.48	1226.77	1204.88
Std. Deviation		88.18	131.25	120.94	130.88
Skewness		-.12	-.01	.04	-.26
Std. Error of Skewness		.62	.26	.62	.26
Kurtosis		-.69	-.45	1.57	-.15
Std. Error of Kurtosis		1.20	.52	1.19	.52
Minimum		1015	846	970	875
Maximum		1320	1408	1469	1495

Note. Time 1 is First Year; Time 2 is Senior Year

Statistical assumptions. Because the intended research design was mixed, all statistical assumptions for analysis of variance (ANOVA) would have been applicable to the analysis. We examined the following statistical assumptions: interval level data, independence of groups, normality, homogeneity of variance, and sphericity. The assumption for interval level data was met, as value-added scores refers to the numerical score on the CLA+ at Times 1 and 2; this variable is continuous, with equal intervals between scores and a true zero value (ratio level). The assumption for independence of groups was met for the between-subjects portion of the statistical test, as students were grouped as only one of the engagement groups: Honors, LP, or Greek Life.

Table 3
Initial Data Distributional Properties for CLA+ Total Scores, Greek Life and Non-Greek Life

		Greek Life Time 1	Non-Greek Life Time 1	Greek Life Time 2	Non- Greek Life Time 2
N	Valid	19	79	19	79
	Missing	0	0	0	0
Mean		1176.54	1100.84	1143.04	1128.05
Std. Deviation		88.18	110.02	130.15	125.80
Skewness		-.12	.69	-.25	.24
Std. Error of Skewness		.62	.52	.27	.52
Kurtosis		-.69	.22	-.29	-.06
Std. Error of Kurtosis		1.20	1.01	.53	1.01
Minimum		1015	960	846	926
Maximum		1320	1366	1408	1403

Note. Time 1 is First Year; Time 2 is Senior Year

Normality, homogeneity of variance, and sphericity. To test the assumption of normality, descriptive statistics were performed to determine the skewness and kurtosis of the dependent variable (CLA+ scores). The values for skewness and kurtosis were converted into standardized scores by dividing the value by the standard error of the value to see whether it was less than an absolute value of 1.96. This was performed for each engagement at both Time 1 and Time 2, and data were determined to be normally distributed for the dependent variable at all Time points. Next, to test the assumption of homogeneity of variance, a Levene's test was performed and found to be statistically nonsignificant across all groups, with Time 1 as the intercept, meeting the assumption for homogeneity of variance. Finally, the test for Sphericity did not yield a result value because there were only two levels of each independent variable (i.e., LP/non-LP, Greek Life/non-Greek Life). Thus, all statistical assumptions were met for the analyses.

Results

We intended to perform 2x2 factorial ANOVAs (mixed between/within-subjects design), with type of engagement (LP only, Greek Life only, and neither) as the between-subjects factors and CLA+ scores from Time 1 (First Year) to Time 2 (Senior Year) as the within-subjects factor. Because the individual engagement group sizes were too small, we could not examine the differences between engagement groups over time; therefore, we have refrained from statements of significance. The overall sample size, however, was large enough to observe the main effect for Time with adequate power, $F(1, 95) = 7.76$, $p = .006$, partial $\eta^2 = .075$. This suggested a potential trend in CLA+ scores over time for all students, regardless of engagement group.

After collecting additional data for the specific engagement groups (Honors, LP, and Greek Life), we will explore how CLA+ scores change over time between groups. At this time, however, we may only observe potential trends for the individual engagement groups descriptively. Table 4 summarizes the mean CLA+ scores for LP and Greek Life at Time 1 and Time 2. Figures 1 and 2 offer a visual representation of the means for each engagement group at Time 1 and Time 2.

Upon observing the graph of mean CLA+ scores over time, the LP and non-LP scores seemed to increase in a similar fashion; the lines from Time 1 to Time 2 appeared almost parallel. The preliminary data suggested little difference in average CLA+ scores between the LP group ($M = 1176.54$, SD = 88.18) and the non-LP group ($M = 1128.48$, $SD = 131.25$) at Time 1 or between the LP group ($M = 1226.77$, $SD = 120.94$) and the non-LP group ($M = 1204.88$, $SD =$

The preliminary data suggested little difference in average CLA+ scores between the LP group and the non-LP group at Time 1 or between the LP group and the non-LP group at Time 2

130.88) at Time 2. LP scores were approximately 4.2% higher at Time 2 than Time 1, and non-LP scores were approximately 6.7% higher at Time 2 than Time 1. Perhaps membership in the LP engagement group would not have an effect on how the CLA+ scores changed over time.

Table 4

Means and Standard Deviations of CLA+ Scores for LP and Greek Life across Time

	Time 1		Time 2	
	M	*SD*	*M*	*SD*
LP	1176.54	88.18	1226.77	120.94
Non-LP	1128.48	131.25	1204.88	130.88
Greek	1100.84	110.02	1128.05	125.80
Non-Greek	1143.04	130.15	1226.96	123.26

At Time 2, the mean scores for the Greek Life group (M = 1128.05, SD = 125.80) were lower than those not in the Greek Life group
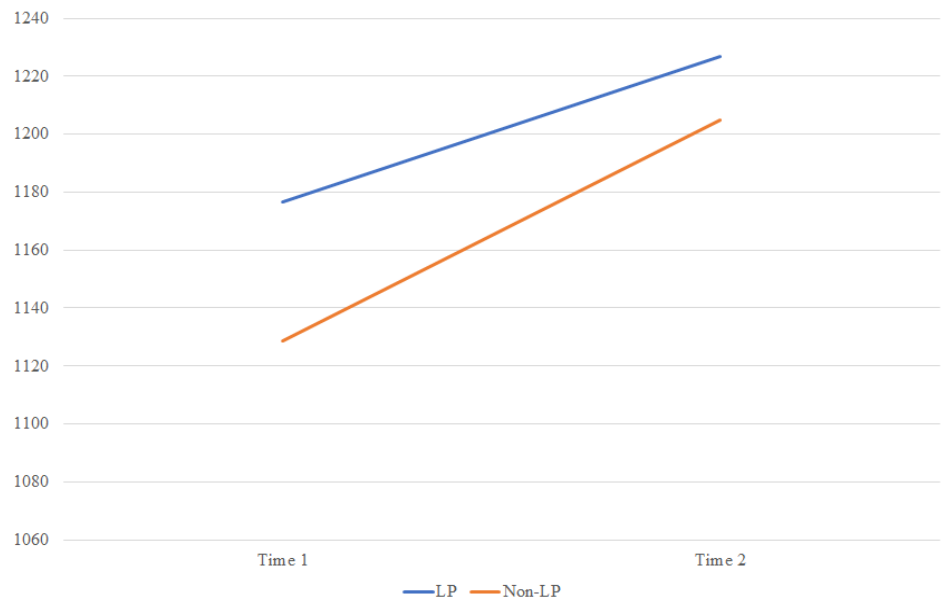
Figure 1. CLA+ scores for LP and Non-LP students from Time 1 to Time 2.

Mean scores on the CLA+ for students who were in the Greek Life group (M = 1100.84, SD = 110.02) did not appear to be different at Time 1 from those not in the Greek Life group (M = 1143.04, SD = 130.15). However, at Time 2, the mean scores for the Greek Life group (M = 1128.05, SD = 125.80) were lower than those not in the Greek Life group (M = 1226.96, SD = 123.26). Greek Life scores were approximately 2.5% higher at Time 2 than Time 1, and non-LP scores were approximately 7.3% higher at Time 2 than Time 1. Perhaps membership in the Greek Life group could have an effect on how the CLA+ scores changed over time. Looking at the two figures comparatively, the mean scores appear to increase differently from Time 1 to Time 2 for Greek Life and non-Greek Life students.
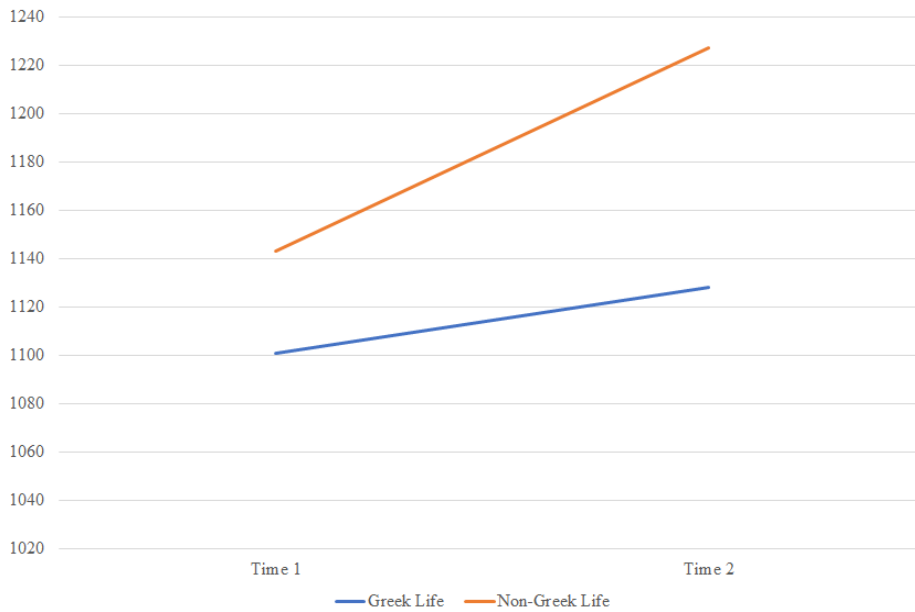
Figure 2. CLA+ scores for Greek Life and Non-Greek Life students from Time 1 to Time 2.

Discussion

This study explored the extent that value-added scores, as measured by performance on the CLA+, varied from First Year to Senior year for students who participated in specific engagement groups on campus. Like many institutions that employ the CLA+, we were interested in the value-add from our institution. Specifically, we were interested in understanding the degree to which specific student engagement activities were related to the success of our students, with the intention of suggesting further support for those practices on campus. The goal was to determine whether engaging in these programs during their undergraduate tenure was contributing to the value-added scores from Time 1 (first year) to Time 2 (senior year) and if scores changed over time differently for different engagement groups.

Overall, students' average scores improved over time, regardless of involvement in an engagement group. That is, scores on the CLA+ improved from First Year to Senior Year for all students. Students in the LP at this institution are expected to engage in program activities throughout their undergraduate studies as a condition of group membership; however, average scores for students in the LP group appeared to increase in a similar manner as those for students in the non-LP group. This finding was surprising because of the well-documented link between behavioral engagement and student success (Kuh et al., 2007; Kuh et al., 2005). To engage academically, a student must invest behavioral, cognitive, and emotional resources (Le, Casillas, Robbins, & Langley, 2005). It is possible, however, that the effect would not be observed due to our definition of value-add (i.e., the change in CLA+ scores from First Year to Senior Year). Perhaps the CLA+ did not capture the type of student gains that could be achieved by participating in the LP due to range restriction. Future research with larger sample sizes will allow us to more clearly observe how scores for LP and non-LP students vary over time.

Similarly, the findings the Greek Life engagement group shed light on potentially unexpected results. While both groups increased their CLA+ scores over time, average CLA+ scores for students in the Greek Life group appeared to increase less drastically than those in the non-Greek Life group. Given that the average Greek Life grade point average (GPA) on our campus is consistently higher than the university average GPA, we expected that active social/emotional engagement on campus might play a role in academic achievement. After all, it has been proposed that emotional engagement precedes, positively influences, and even facilitates cognitive engagement (Fredricks et al., 2004) and that engagement research has demonstrated

Overall, students' average scores improved over time, regardless of involvement in an engagement group

that students who are involved in their academic communities may have positive academic outcomes (Zhao & Kuh, 2004; Mouzakis, 2017). As we gather more longitudinal data in the coming years, we plan to explore this relationship further.

According to a 2009 College Senior Survey reported at the Higher Education Research Institute (Franke, Ruiz, Sharkness, DeAngelo, & Pryor, 2010), 17.5% of graduating college seniors report having joined a fraternity or sorority. As discussed in a recent article by Even and Smith (2018), these national organizations often outline goals for their membership such as the "intellectual development of members, the cultivation of leadership skills, and citizenship" (Even & Smith, 2018, pp.1). Why then, with these national membership requirements and with the observed academic performance of our Greek Life students, do our results demonstrate that Greek Life participants scored significantly lower on the CLA+ than non-Greek Life participants during their senior year when their first-year scores were not significantly different? Perhaps the low-stakes nature of taking the CLA+ at this institution played a role. Finney, Sundre, Swain, and Williams (2016) noted that test-taking motivation had an effect on estimates of value-add, in that a student's motivation influenced their performance on a test after taking other variables into account (Finney, Sundre, Swain, & Williams, 2016). This phenomenon may apply to the present study because completing the CLA+ was not a requirement at this institution, and the results did not become part of students' academic records.

It is important to continue specifying the type of student engagement being examined, rather than grouping participation in activities under the umbrella of "student engagement"

Study Limitations

The specific engagement groups were small in size, which presented challenges in capturing interaction effects and between group effects. This sample represented the first grouping of longitudinal data at our institution, and we hope to obtain larger sample sizes in the future to allow for more powerful analyses. Also, students were not required to take the CLA+ as part of their academic studies. Motivation (or lack thereof) to perform well on the test may have influenced the outcome at either test administration (Williams, 2015). Although the sample of students who were recruited to take the CLA+ was random and generally reflected the composition of the student population at this institution, the test was still considered to be low-stakes.

The current study was performed at an institution using existing data and groups for engagement, defining types of engagement and value-add in a very specific way that may not be applicable to all institutions. The method we used to determine value-add may not have fully captured the particular effects of these specific student engagement activities; however, in explicitly describing the type of student engagement being explored in this study, it may contribute to the literature by examining how different types of engagement play a role in institutional value-add.

Conclusions and Future Directions

Depending on how one defines engagement (cognitive, behavioral, emotional, etc.), student participation in such activities may have a different influence on institutional value-add. It is important to continue specifying the type of student engagement being examined, rather than grouping participation in activities under the umbrella of "student engagement." Doing so may help to improve our understanding of how high-impact practices support student success. Other activities that may be interesting to explore as subtypes of student engagement might include athletics, student government, service organizations, and cultural clubs.

Due to the very small sample size of the Honors Program group, we were not able to explore the potential effect of cognitive engagement in this study. As our sample of longitudinal data increases over the coming years, we would like to examine how CLA+ scores vary over time for Honors versus non-Honors students. Further, we would like to continue exploring the potential role of participating in LP and Greek Life on student achievement. Upon gathering enough data for adequate sample sizes of groups, we plan to perform a 3x2 mixed factorial analysis of variance (ANOVA) to explore how student performance on the CLA+ varies over time for different engagement groups.

We were initially interested in triangulating these findings with NSSE data; however, we did not have an adequate number of matched cases, as we cannot control which students complete the NSSE or are randomly selected to take the CLA+. Because the manner in which value-added is measured by an institution may not fully reflect gains in the student experience, it may be helpful to take student perceptions into consideration. Also, our method of calculating value-added scores (i.e., using longitudinal data to see the change in scores from Time 1 to Time 2) is different than that of CAE's calculation for value-add (which takes parental level of education into account). Therefore, we may wish to perform a comparative analysis of methods for calculating institutional value-add, noting the potential difference and implications of each method. Finally, in examining the data and results we would like to further explore the impact of students entering the university with college credit (as defined by the number of advanced placement/dual-enrollment credits) on institutional value-add.

References

About CAE. (2018). Retrieved from https://cae.org/about-cae/history/

Axelson, R. D., & Flick, A. (2011). Defining student engagement. Change, 43(1), 38. doi:10.1080/00091383.2011.533096

Douglass, J. A., Thomson, G., & Zhao, C. (2012). The learning outcomes race: The value of self-reported gains in large research universities. Higher Education, 64(3), 317–335. doi:10.1007/s10734-011-9496-x

Even, W. E., & Smith, A. C. (2018). Greek Life, Academics, and Earnings. IZA Institute of Labor Economics, IZA DP No. 11841. doi:10.2139/ssrn.3257025

Finney, S. J., Sundre, D. L., Swain, M. S., & Williams, L. M. (2016). The validity of value-added estimates from low-stakes testing contexts: The impact of change in test-taking motivation and test consequences. Educational Assessment, 21(1), 60–87. doi:10.1080/10627197.2015.1127753

Franke, R., Ruiz, S., Sharkness, J., DeAngelo, L., & Pryor, J. (2010). Findings from the 2009 administration of the College Senior Survey (CSS): National aggregates. Higher Education Research Institute at the University of California Los Angeles, Los Angeles, CA.

Fredricks, J.A., Blumenfeld, P.C., & Paris, A.H. (2004). School engagement: Potential of the concept, state of the evidence. Review of Educational Research, 74(1), 59–109. doi:10.3102/00346543074001059

Handelsman, M. M., Briggs, W. L., Sullivan, N., & Towler, A. (2005). A Measure of College Student Course Engagement. The Journal of Educational Research, 98(3), 184–191. doi:10.3200/JOER.98.3.184-192

Klein, S., Benjamin, R., Shavelson, R., & Bolus, R. (2007). The collegiate learning assessment: Facts and fantasies. Evaluation Review, 31(5), 415–439. doi:10.1177/0193841x07303318

Kuh, G. D., Cruce, T., Shoup, R., Kinzie, J., & Gonyea, R. M. (2007, April). Unmasking the effects of student engagement on college grades and persistence. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. doi:10.1353/jhe.0.0019

Kuh, G.D., Kinzie, J., Schuh, J.H., & Whitt, E.J. (2005). Never let it rest. Change, 37(4), 44–51. doi:10.3200/CHNG.37.4.44-51

Lester, D. (2013). A review of the student engagement literature. Focus on Colleges, Universities, and Schools, 7(1), 1–8.

Le, H., Casillas, A., Robbins, S. B., & Langley, R. (2005). Motivational and skills, social, and self-management predictors of college outcomes: Constructing the student readiness inventory. Educational and Psychological Measurement, 65(3), 482–508. doi:10.1177/0013164404272493

Lodge, J. M., & Bonsanquet, A. (2014). Evaluating quality learning in higher education: Re-examining the evidence. Quality in Higher Education, 20(1), 3. doi:10.1080/13538322.2013.849787

Mouzakis, K. (2017). Academic and social engagement in university students: Exploring individual differences and relations with personality and daily activities (Order No. 10618229). Available from ProQuest Dissertations & Theses Global. (2009467037).

The College Board (2019). Trends in college pricing: Highlights. Retrieved from https://research.collegeboard.org/trends/college-pricing/highlights

U.S. Department of Education, National Center for Education Statistics. (2019). Fast facts: Tuition costs of colleges and universities. Retrieved from https://nces.ed.gov/fastfacts/display.asp?id=76

Williams, L. M. (2015). The effect of examinee motivation on value-added estimates (Order No. 3702699). Available from ProQuest Central; ProQuest Dissertations & Theses Global. (1685390194). Retrieved from https://0-search-proquest-com.read.cnu.edu/docview/1685390194?accountid=10100

Zhao, C-M., & Kuh, G. D. (2004). Adding value: Learning communities and student engagement. Research in Higher Education, 45, 115–138. doi:10.1023/b:rihe.0000015692.88534.de

Abstract

In higher education, many barriers exist to efficient outcomes assessment, one of which is the culture that exists within institutions and the perceptions of the faculty who are at the forefront of assessing student learning. It is often viewed as a "culture war" (Baas, Rhoads, & Thomas, 2016) as faculty remain unconvinced that what the assessment movement sets out to achieve is actually attainable. For this to gain traction and achieve its intended outcomes, the process directives and tasks must be amenable to the individuals who provide most of the data, the faculty. Since metrics are of utmost importance in the assessment world, appraisal of the process at a small medical college most appropriately utilized Q methodology, which provides the basis for the scientific study of subjectivity. Thirty-four statements were derived through interviews with the faculty and the Q-sorts were completed by 14 faculty and 4 staff from the same body. Two well-defined factors emerged. One group of participants believed in the movement, along with the benefits of assessment. The other group, surprisingly, was not the antithesis, but rather expressed concerns about the lack of time and resources dedicated to the data gathering as well as the possible punitive uses of the results. The factors that emerged could play a vital role in the adjustment and improvement of the process.

AUTHORS

Viveka Jenks & Abby Green

Kent State University

Appraisal of the Curricular Assessment Process

Accreditation agencies in higher education have been increasingly emphasizing curricular assessment since the acceleration of the assessment movement in 1985 (Banta, 2002). Institutions can no longer make decisions based on hunches, intuition, and informal data. For approval by the agencies, institutions need to have concrete data to support the assessment of student learning. Accrediting agencies as well as prospective and current stakeholders have a vested interest in ensuring that colleges are accountable for not only delivering on the student learning outcomes promised, but also for monitoring assessment closely enough to make small or large changes when outcomes are not favorable (Anderson, 2004). This brings to light the discrepancy and difference in utility between internal and external purposes of assessment. The internal purpose of assessment is curricular improvement whereas the external purpose is accountability. The challenge lies in providing enough evidence to external accrediting bodies without neglecting the internal purpose of quality improvement. However, despite all of the complications and nuances, student learning remains the central purpose of the assessment process.

Assessment has been the topic of many heated discussions since it began. There have always been the advocates of outcomes assessment who feel it is essential to measure the effectiveness of all outcomes in higher education. However, equally as vocal are the higher education professionals who are the so-called "skeptics." According to the literature, the range of concerns for such professionals encompasses the political nature of assessment, the lack of meaningful faculty input and contribution, and the validity of the measure of outcomes commonly employed (Baas et al., 2016). In addition to assessment being described as a culture war among faculty, they are often hesitant to move beyond their typical frames of reference to assess higher-order learning outcomes, and the procedures can seem complex and unrelated to their work (Anderson, 2004).

CORRESPONDENCE

Email:
vjenks@kent.edu

A podiatric medical college recently proposed a curricular revision that was met with resistance from the accreditation agency until a full curricular assessment plan was established. The resistance was due to the lack of specific data that pointed directly to the impetus for the change. As a response, a curricular assessment plan was established and rolled out to the faculty in 2017. The process involved mapping course objectives to course goals and then to program objectives. The course objectives reflect what the students learn during each encounter; for example, "The student will be able to describe ageism and some of the associated myths." Each course objective should relate to a course goal such as "Appreciate the health care concerns associated with the aging population." Program objectives are the overall competencies that the college has created that each student should possess before graduation. The program objective that fits with the example stated here is "Recognize the important determinants of changing health that contribute to the development and/or continuation of illnesses." To accomplish learning assessment at the most basic level (objectives), faculty were expected to examine their individual course exams or observational assessments and demonstrate how each of the proposed course objectives were measured. The proposed hierarchy of tagging exam items is shown in Figure 1.

The resistance was due to the lack of specific data that pointed directly to the impetus for the change
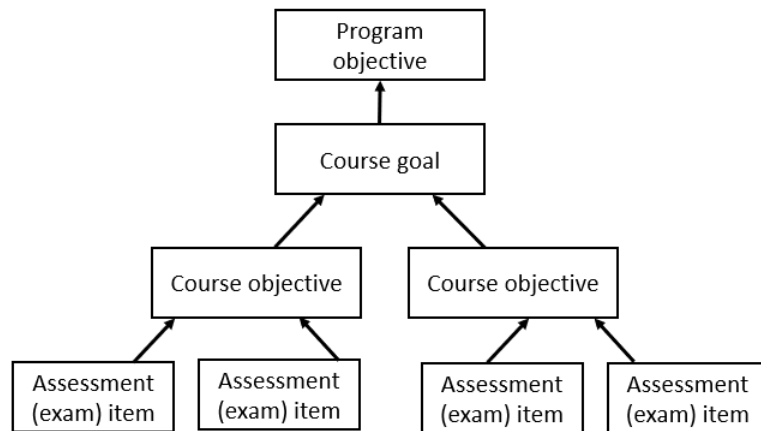

Figure 1. Hierarchy of tagging assessment (exam) items

For the most part, the didactic lecture courses were the least complex to assess, as faculty were simply required to tag their exam items (primarily multiple choice questions) to their course objectives, course goals and the college's program objectives using a secure assessment software product (Examsoft, n.d.). Administrative staff mapped the items to institutional goals (such as those stated by the university). The process was an initial time investment for each faculty member, as each question needed to be closely examined to identify the appropriate tags. Care was taken to train faculty on the basic directions, and during follow up meetings it was explained that, if this process were done efficiently, it would require minimal maintenance. Complete tagging will only need to occur if entirely new questions were added on subsequent exams. Once all exam item data (tags) had been gathered and entered into the assessment software, reports were run that aggregated student achievement by category (college program objective, course goal, and course objective). From these reports, an adequate assessment of the strengths and challenges of the curriculum could be made.

Through informal conversations, both positive and negative opinions were expressed throughout the college about this new endeavor. Without full faculty buy-in, however, the assessment process would not be sustainable. To achieve the intended outcome of implementing the assessment process, empirical data on faculty perception had to be obtained. This was accomplished through research utilizing Q methodology. This method was appropriate for revealing the subjective perspectives of the faculty and staff who provide the data that is relied upon for accurate curricular assessment. Q methodology contains elements of both qualitative and quantitative research, which will provide administrators of the program with adequate insight into what changes are needed in the program to make it sustainable. The intent of this research was to identify the real concerns of faculty about the assessment process, and then to use the data

for improvement to ensure the program's ongoing utility. Q methodology allows the researcher to understand individual subjectivity (or groupings of opinions) without aggregating the data and thus diluting the opinions. The opinions of relevant staff members were also solicited based on how they perceived the faculty felt about the process. Relevant staff members are defined as the staff who work closely with the faculty responsible for tagging exam items, and thus their perspective was valuable.

Development of Concourse and Q Sample

All studies that involve Q methodology have data rooted in communication concourse (Stephenson, 1978), which has been characterized by McKeown and Thomas (2013) as "often ambiguous, utterly subjective, semantically imprecise, yet wholly natural condition of much of human communication" (p. 17). Concourse data can be derived from anywhere subjective communication resides such as internet discussion boards, formal and informal interviews, open-ended survey responses, and polls. In this study, the concourse data collection took the form of semi-structured interviews among the faculty and relevant staff members of a podiatric medical college in northeast Ohio. Prior to participant recruitment and interviews, the study was reviewed and approved by the college's institutional review board. All participants were required to sign a consent form prior to being interviewed. Donner (2001) suggests asking "umbrella questions that allow multiple possible answers" (p. 26) so as to generate a depth and breadth of responses. Examples of questions that were asked of the faculty and staff were as follows:

1. What are your thoughts about the current curricular assessment process?
2. Do you have any suggestions for improvement?
3. What is your reaction to comments that the assessment process can be used for punitive purposes? The perception that it is busy work, etc.? Please elaborate.

The subjective data derived from the interviews were reduced to 34 individual statements that were randomly numbered and printed on individual pieces of paper. Care was taken to include approximately equal numbers of positive and negative statements as well as to maintain the representativeness of the concourse. The entire set of statements is shown in the Appendix.

Sorting of the Q Sample

Participation in the Q-sort was voluntary. Additionally, the completed Q-sort grids were anonymous to the primary investigator (unless the participant revealed his or her identity) to ensure complete honesty, as the primary investigator was the lead staff person for curricular assessment purposes. Fourteen faculty members and four staff members chose to participate in the Q-sort. Gender or division information (such as preclinical science faculty, podiatric medicine faculty, or surgery faculty) were not recorded due to the small size of the participant pool and the likelihood of identities being revealed with such information. Participants were provided with detailed instructions about how to do the Q-sort via a PowerPoint presentation as well as a printed instruction sheet during a faculty meeting. Each participant received an envelope with the 34 statement cards, the scoring sheet, and a consent form if they had not participated in the interview phase of the study (participation in the Q-sort was not dependent on prior participation in the interview phase). Faculty were asked to sort the 34 statements under the following condition of instruction: "Please use the cards to describe your experience with and opinion of the assessment process by ranking the appended statements from 'most agree' (+4) to 'most disagree' (-4)." Similarly, key staff were provided the following condition of instruction using the same Q-sort statements and scoring grid: "Please use the cards to describe how you feel the faculty, in general, feel about their experience (based on your interactions with them) and opinion of the assessment process by ranking the appended statement for 'most agree' (+4) to 'most disagree' (-4)." The Q-sort grid is usually a quasi-normal distribution scale with equally weighted positive and negative sides, as shown in Figure 2.

Most faculty completed their Q-sorts within 20 minutes and returned their score sheet, whereas other participants submitted the score sheet later that week. The primary investigator (VJ) left the room while the participants sorted the cards. The coinvestigator (AG), who also
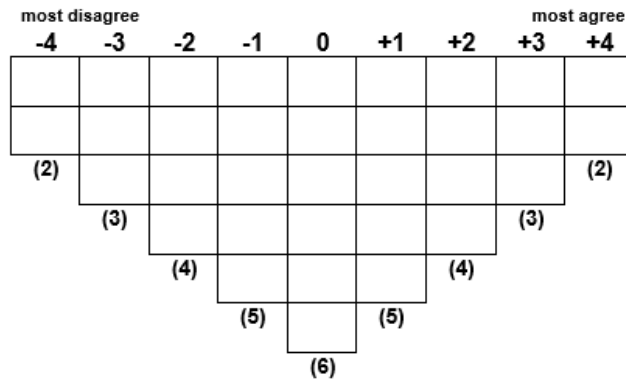
Q methodology contains elements of both qualitative and quantitative research, which will provide administrators of the program with adequate insight into what changes are needed in the program to make it sustainable

Figure 2. Sorting grid for this study

The analysis resulted in two factors, or schools of thought, among the participants who are actively involved in the assessment process

conducted the interviews, remained in the room to observe participant interaction with the statements and was available to answer any questions. Participants were encouraged to comment on the Q-sort process on the flip side of their score sheets, answering the question, "Please feel free to comment on anything (the Q-sort process itself, any other comments/feelings about the assessment process in addition to, or that wasn't covered by the statements, etc.)." Other qualitative information was recorded by the coinvestigator while observing and interacting with the participants as they were conducting the Q-sort, such as inquiries about possible hesitations, changes of mind, and frustrations.

Analysis

A total of 18 participants provided their opinions by sorting the 34 statements from the Q-sample. Four of the participants were staff who worked closely with the faculty on the assessment process. The faculty participants (14) represented 67% of the total full-time faculty body. The staff participants represented 57% of all staff who work closely with faculty and their exams (excluding the primary investigator).

Using the PQMethod program (Schmolck, 2014), the 18 Q-sorts resulted in an 18 x 18 correlation matrix, which was then factor analyzed using the principal components method. The factors were then rotated by varimax. The analysis resulted in two factors, or schools of thought, among the participants who are actively involved in the assessment process. Factor scores (from +4 to –4) were then estimated for each of the 34 statements within each of the factors (see Appendix). Each sort loaded on one or the other factor. There were no sorts that loaded on both factors (meaning that their views possess elements of both), and there were no sorts that were not defined by a factor (see table 1). The correlation between the factors was -0.1708, indicating two distinct points of view.

Table 1. Factor matrix with an X indicating a defining sort

QSORT	1	2	QSORT	1	2
1	0.6535X	0.2859	10	0.7592X	-0.276
2	0.7448X	-0.3884	11	0.3112	0.7758X
3	0.7108X	-0.1772	12	0.0084	0.6362X
4	0.7267X	-0.4107	13	0.4364X	0.4101
5	0.6138X	0.2201	14	0.8427X	-0.2323
6	0.6397X	0.2054	15	-0.0735	0.5557X
7	-0.0045	0.7453X	16	-0.1437	0.7493X
8	0.3617	0.4831X	17	-0.0092	0.4101X
9	0.6551X	0.2754	18	-0.4819	0.7604X

Results

Factor 1: We think assessment is helpful!

Overall, participants that loaded on factor 1 had a positive opinion about the assessment process. Considering that the factor consisted primarily of faculty members (9 of the 10 defining Q-sorts), the results suggest that faculty sentiment is mainly in line with factor 1. However, this would have to be confirmed using a more comprehensive survey. The positive perspective of factor 1, especially when contrasted with factor 2, was seen in the following statements (scores to the right for factors 1 and 2, respectively):

12. Assessment data is important in order to reflect properly upon that data and make improvements.	4	0
29. Everyone should view assessment as part of their course as opposed to something external.	4	0
1. Assessing our course is the only way we're going to know where we've been and where we're going.	3	1
31. Assessment makes you more efficient as a faculty member and saves you time in the long run.	3	-2
8. I think this is a genuine attempt to address complaints and critiques on assessments.	3	-2

The two highest-ranked statements (#12 and #29) highlighted the importance of assessment to factor 1 and the use of data in supporting evidence of student learning. The three statements that ranked second highest supported this by underscoring the use of assessment in making a faculty member more efficient (#31), knowledgeable about learners (#1), and attuned to the students' experience (#8).

Examination of those statements with which factor 1 "most disagree" helped to clarify the perspective that this group of participants espoused:

2. This is all bullsh*t.	-4	-1
26. In some cases, the reflection forms infringe on our academic freedom.	-4	0
25. The Curriculum Assessment Committee needs to be eliminated.	-3	-1

Not surprisingly, factor 1 rejected the most evocative statement (#2) that suggested that the assessment program is nonsense and disagreed strongly that the program infringed on their academic freedom (#26). In disagreeing with statement 25, factor 1 felt that the Curriculum Assessment committee has its place in the school and need not be eliminated. Participants that loaded on factor 1 were comfortable meeting with their division head to discuss their reflection forms and did not view the completion of this form and the reflection meeting as intended for punitive purposes.

Five of the 10 score sheets for factor 1 participants included handwritten comments, one in particular highlighting the heart of factor 1 and attesting to the operantcy of the factor scores as well as to the utility of the Q-sorting process itself: "I think the general trend, using a more innovative and well thought-out assessment process, is long overdue (including [the] Q-sort process). Assessment processes are never perfect, but improvement in functionality is long overdue." Another written comment also reaffirmed the Q factor by indicating that the negative statements were perplexing given that the participant appreciated that assessment is a "useful and meaningful way" to alert faculty as to how well the students are learning the content.

Factor 2: We need more time for assessment and assurance of nonpunitive use!

At first glance, the highest-ranking statements for factor 2 did not emerge as a complete antithesis of factor 1 but rather expressed a separate concern––namely, with the expenditure of time and effort (scores to the right for factors 1 and 2, respectively):

The two highest-ranked statements (#12 and #29) highlighted the importance of assessment to factor 1 and the use of data in supporting evidence of student learning

6. If our exam questions need to be tagged on ExamSoft, someone needs to give us more time to do it. We do not have time.	0	4
9. It is a challenge to get all the tagging done because of how time consuming the assessment process is.	-1	4
4. The reflection forms can be used for punitive purposes.	-3	3
24. I think the assessment methods require a lot of effort, but I am not sold on the data we receive from them. I don't know if the information we receive from tagging is valuable yet.	0	3

The two factors emerged as orthogonal (hence, uncorrelated with one another overall), but they differed with one another on specific statements while agreeing with one another on other statements. If they were truly antithetical, there would be only one bipolar factor. The most highly ranked statements (#6 and #9) as well as one of the next-highest ranked statements (#24) expressed concerns about tagging being a time-consuming process. The score for statement 4 introduced an important political consideration––that the reflection forms might be used for punitive purposes––and this was echoed in other statements (e.g., #33, see Appendix). As these scores indicate, factors 1 and 2 are divided about whether the administration can be trusted, and this issue is apparently critical when it comes to supporting or opposing the assessment process.

many of the nonfaculty participants identified factor 2 rather than factor 1 as the viewpoint of the faculty as a whole, a misperception...

The statements with which factor 2 disagreed reinforced much of what has already been noted:

30. I wish this assessment process was something we did years ago.	0	-4
32. There are people who have plenty of time to tag questions, but just don't want to.	-1	-4
14. Great strides are being made in the assessment process. We are constantly improving and evolving.	1	-3
23. I believe my division head protects me from the reflection forms being used as punitive purposes.	1	-3

Limited time again emerged as a concern for the factor (#32), as did the potential for punitiveness (#23). Factor 2 questioned the very legitimacy of the assessment process and therefore did not consider that great strides had been made in it (#14), nor did these individuals wish that it was done years ago (#30).

One surprising point is that factor 2 consisted of only 62.5% of the faculty participants (5 of 8), the remainder being staff members who had submitted Q-sorts of their perceptions of the faculty. That is, many of the nonfaculty participants identified factor 2 rather than factor 1 as the viewpoint of the faculty as a whole, a misperception that could either be due to the staff having been disproportionately targeted for expressions of discontent from faculty associated with factor 2, and/or due to some of the more disgruntled faculty disguising their sentiment behind a more positive expression of support for the assessment process.

Despite demonstrable differences between the two groups, there were a few statements that achieved a degree of consensus between factors 1 and 2. In many instances, however, the points of agreement appeared close to the zero point of neutrality for both of the factors (e.g., statements 10 and 22 in the Appendix). One statement achieved a mild level of cross-factor agreement, however, and another acquired a relatively high level of negative saliency:

15. The statement that assessments are just busy work is partially true. Sometimes the methods we use to accommodate the accrediting bodies is looked at as simply a check-off box. I fear whether we are ever really achieving our set goal of making a class better.	1	2

18. I think that each faculty member should have to meet with a dean to   -3   -3
    discuss their course and student evaluation rather than a division head.
    want to.

      The mildly agreeable concern that the assessment process is just busy work and of doubtful utility (statement 15) seemed to be a popular viewpoint among faculty. This has also been observed in other assessment research where faculty are unconvinced that the aspirations of the assessment movement are achievable (Baas et al., 2016). Opposition to the view that faculty members should meet with the dean rather than division head (#18) may be less a matter of congruence than a function of differences in perspective––factor 1 believing assessment to be more effective when determined locally, and factor 2 being more concerned about assessment discussions between parties of unequal authority.

Conclusions and Recommendations

      The views of the faculty regarding the assessment process are promising, realistic, thought provoking, and helpful for the administrators involved in coordinating the assessment plans and directives. Overall, the program shows promise, with a few concerns for time investment and fear of punitive consequences. The staff seemed to reflect more of the concerns the faculty had rather than the aspects with which they were satisfied. This is expected as faculty may project their frustrations more than their satisfactions about the assessment process to their staff, but it does not mean they were not satisfied with the program or found it beneficial. The results of this study will be used to inform changes in the assessment program as well as improvements to encourage more faculty involvement. It is evident from the data that the greatest concern is time, and it would be a good idea to incorporate release time within the academic year in the form of semiannual retreats to allow for faculty to tag their questions and reflect on their course assessment data. The other concern to be aware of is that the process may be a waste of time if the course data retrieved from the assessment process is not actually used to drive and inform changes in the curriculum, both for internal and external purposes. Factor 2's disagreement with statement #14 indicates the perception that the assessment program is not improving. This viewpoint is valuable and provides avenues for further research once the necessary alterations have been made to the assessment program.

It is important to understand that the participants in a Q study do not represent a population, nor should the results be generalized to the population

      It is important to understand that the participants in a Q study do not represent a population, nor should the results be generalized to the population. The aim of Q methodology is not quantitative generalizability, but rather to determine how subjectivity is clustered among participants (Shemmings & Ellingsen, 2012). However, a clear advantage of the Q technique is the ability to provide the researcher with a better understanding of groupings that exist in a population so to fuel further probing if necessary (Brown, 2002). In future Q-sorts it would be optimal for the primary researcher to be present during the participants' Q-sort in order to have more of an understanding of the sorters' thought processes while deliberating with the statements. In this particular study it was advised that the primary investigator not be present as this individual is the creator of the newly implemented assessment program, and their presence may have created an environment in which faculty were not comfortable being completely honest if they disagreed or disliked the program. However, as the assessment process is further refined considering the results of this research study, trust in the investigator and willingness on the part of the faculty to talk directly with the investigator may both increase in subsequent studies.

      The study did include several shortcomings, especially since not all faculty participated. Widening the application of the Q-sort may have revealed yet another factor that was not exposed in this study. Future research includes using a similar Q-sort with some statements that are unique to the school's structure eliminated and a greater sample size of faculty and staff who share a similar role in the use of the assessment software. This will provide ideas on which parts of the assessment process faculty are satisfied with and those they would like to be improved upon. It should be noted that the expletive used in statement #2 was an actual quote derived from the interviews and was retained for the Q-statement cards to determine if other faculty felt just as strongly about the assessment process.

Appendix.
Factor Scores for All Statements

Statement	Factor 1 Array	Factor 2 Array
1. Assessing our course is the only way we're going to know where we've been and where we're going.	3	1
2. This is all bullsh*t.	-4	-1
3. Tagging questions on ExamSoft is just busy work.	-2	1
4. The reflection forms can be used for punitive purposes.	-3	3
5. Depending on your division head, the reflection form could be a really good tool.	1	2
6. If our exam questions need to be tagged on ExamSoft, someone needs to give us more time to do it. We do not have time.	0	4
7. CPME doesn't want more data collection. They don't want to look at our individual questions, so I don't know why we need to tag questions.	-2	1
8. I think this is a genuine attempt to address complaints and critiques on assessments.	3	-2
9. It is a challenge to get all the tagging done because of how time consuming the assessment process is.	-1	4
10. I think there are attempts to educate faculty on the assessment process, but I think it needs to be continued and reinforced so that we don't find it as intimidating.	1	0
11. I completely disagree with the statement that assessments are just busy work and have no effect on instruction. I find that statement frustrating.	0	-2
12. Assessment data is important in order to reflect properly upon that data and make improvements.	4	0
13. Reflection forms should be done more honestly. Reflection should be more open-minded.	2	0
14. Great strides are being made in the assessment process. We are constantly improving and evolving.	1	-3
15. The statement that assessments are just busy work is partially true. Sometimes the methods we use to accommodate the accrediting bodies is looked at as simply a check off box. I fear whether we are ever really achieving our set goal of making a class better.	1	2
16. It is a real issue when a faculty member must meet with a division head to reflect on a course the division head knows nothing about.	-1	3
17. We need to go back to the old reflection forms.	-2	-1
18. I think that each faculty member should have to meet with a Dean to discuss their course and student evaluation rather than a division head.	-3	-3
19. We need to put the emphasis on the major goals and program objectives rather than laser thin subcategories.	0	2
20. I receive most of my information from the student evaluations rather than the reflection forms.	0	1

21. I am concerned about how the reflection forms are written. The way they are set up, it doesn't encourage us to answer honestly.	-1	1
22. The assessment process is not perfect, but it is fine the way it is.	-1	-1
23. I believe my division head protects me from the reflection forms being used as punitive purposes.	1	-3
24. I think the assessment methods require a lot of effort, but I am not sold on the data we receive from them. I don't know if the information we receive from tagging is valuable yet.	0	3
25. The Curriculum Assessment Committee needs eliminated.	-3	-1
26. In some cases, the reflection forms infringe on our academic freedom.	-4	0
27. We need to invest in curriculum mapping software and have our entire curriculum mapped before we start doing deeper overall curricular assessment.	2	-2
28. Our goal is to improve the educational process. We do that through assessment.	2	0
29. Everyone should view assessment as part of their course as opposed to something external.	4	0
30. I wish this assessment process was something we did years ago	0	-4
31. Assessment makes you more efficient as a faculty member and saves you time in the long run.	3	-2
32. There are people who have plenty of time to tag questions, but just don't want to.	-1	-4
33. Assessment is not done for students, but for administrators. Not for faculty, but to faculty. Not for program improvement, but for compliance monitoring.	-2	2
34. I like assessment because it encourages faculty members to think more carefully about what they do, how we do it, and why we do it that way.	2	-1

References

Anderson, J. (2004). An institutional commitment to assessment and accountability. In P. Hernon & R. Dugan, Outcomes Assessment in Higher Education. Views and Perspectives. Westport, CT: Libraries Unlimited.

Baas, L., Rhoads, J. C., & Thomas, D. B. (2016). Are quests for a "culture of assessment" mired in a "culture war" over assessment? A Q-methodological inquiry. SAGE OPEN, 6(1). https://doi.org/10.1177/2158244015623591

Banta, T. (2002). Building a scholarship of assessment (1st ed.). San Francisco, CA: Jossey-Bass, A Wiley Brand.

Brown, S. R. (2002). Q Technique and Questionnaires. Operant Subjectivity, 25, 117–126.

Donner, J. (2001). Using Q-sorts in participatory processes: An introduction to the methodology. In Paper No. 36. Social analysis: Selected tools and techniques (pp. 23–49). Washington, DC: Social Development Department, The World Bank.

Examsoft. (n.d.). Dallas, TX: Examsoft Worldwide, Inc.

McKeown, B., & Thomas, D. B. (2013). Q Methodology. SAGE Publications.

Schmolck, P. (2014). PQMethod (Version 2.35). Retrieved from http://schmolck.userweb.mwn.de/qmethod/#PQMethod

Shemmings, D., & Ellingsen, I. T. (2012). Using Q methodology in qualitative interviews. The SAGE handbook of interview research: The complexity of the craft. Thousand Oaks: SAGE Publications, 415–27.

Stephenson, W. (1978). Concourse theory of communication. Communication, 3, 21–40.

Abstract

Academic assessment faces many challenges, one of which is low participation on the part of faculty members (e.g. Smith, 2005). Here we present a case study from our assessment work in higher education. In addition to updating the instructions we created assessment documents for Hogwart's School of Witchcraft and Wizardry as a template for faculty to use. The data suggest that presenting assessment to faculty in a humorous way, and geared toward faculty concerns, can both increase participation and the quality of the participation.

AUTHORS

Julia J. C. Blau
Central Connecticut State University

Eric T. Blau
Portland, CT

The Use of Humor in Assessment: A Case Study

I joined the Academic Assessment Committee (AAC)[1] at one of my previous institutions for the same reason that everyone did: because I was volun-told to. The dean noted that, since my husband worked in our administration in institutional assessment, I must be well suited to the position. At the time, all I knew about assessment was that the word left a bad taste in everyone's mouth and that the entire process was seen as a bureaucratic waste of time. The reality was much worse.

At the time of my joining only seven academic programs were fully participating in Middle States Commission on Higher Education (MSCHE) assessment requirements for accreditation. There were seven members on the AAC (We encourage the reader to draw their own conclusions about that coincidence). Without any incoming plans or annual reports to review, the work of our small committee was largely drafting increasingly irritated emails begging departments to participate.

Eventually, I surmised that the resistance to participation was likely due to two factors. First, the faculty and the administration had very different views on the purpose and usefulness of assessment—a concern that will, no doubt, sound familiar. Though it is easy to dismiss faculty disengagement as a byproduct of laziness (Smith, 2005), the reality is probably more complex. My observations are in line with Ewell's (2002; see also Banta, 2002) observation that faculty find assessment redundant and invasive: As part of their annual renewal process (and just being good at their jobs more generally) they already review their teaching methods and attempt to find better pedagogy, so AAC's insistence on external assessment struck faculty members as questioning their facility with self-review (Emil & Cress, 2014). This was a source of particular anxiety for pre-tenure faculty, who worried that the assessment procedures would be used against them when it came time for renewal, tenure, or promotion. Moreover, if the faculty were already instituting the necessary pedagogical changes to increase student learning, what was the purpose of the assessment cycles beyond keeping the administration at bay? It was little wonder

CORRESPONDENCE

E-mail: blau@ccsu.edu

[1] School, committee, and office names changed to protect the innocent.

that so few departments participated. As Driscoll and De Noriega (2006) pointed out, resistance dramatically decreases when the usefulness of assessment is linked to the issues that matter to professors: increasing student learning and improving the craft of teaching.

The second problem was the instructions AAC used to assist departments in creating their assessment documents. Beyond being nearly incomprehensible and filled with assessment jargon, they did nothing to combat the misconceptions held by the faculty about the purpose and usefulness of assessment. It was clear to me that the answer was to rewrite the instructions document in a humorous and accessible way that would painlessly explain the purpose and the benefit of assessment. In keeping with Emil and Cress's (2014) observation that faculty's belief in the usefulness of assessment is directly proportional to the amount of effort they are willing to expend on it, I hypothesized that the release of such a document would increase both the participation and the quality of participation.

Hogwarts to the Rescue

With the help of my assessment-savvy husband, I created a series of documents as if I was the chair of the Defense Against the Dark Arts Department at Hogwarts School of Witchcraft and Wizardry (Rowling, 1997-2007). These documents were then integrated into an entirely new set of instructions written with the express goal of clarity and transparency. We took care to address every situation that might arise at our institution. For example, one of our departments relied on an external test as an assessment measure so we worked in the Ordinary Wizarding Level test—an external test that already existed in the universe of Hogwarts.

We also addressed the primary issue: explaining how assessment might be useful to the departments. From our view, academics do not need help being explicit about changes that they can make to help improve student learning outcomes. Explicit self-reflection is already included in the renewal, tenure, and promotion process (not to mention department meetings, countless informal discussions, and internal reflection). What the assessment process can uniquely offer academics is communication. Faculty do not get many opportunities to communicate their needs in a way that is compelling to administrators (Williams, 2013). When the assessment process is focused only on what individual professors or departments can do to improve student learning outcomes, it is unsurprising that it feels redundant and useless. Our instructions shifted the focus of the final "closing the loop" section from exclusively internal solutions (such as reorganizing the presentation of material) to including external ones (such as decreasing class size or changing the course prerequisites).[2]

Method

Corpus and Design

In order to complete a within-subjects design we used records from those departments for which recent historical data were available (n = 28). Some programs were discontinued, and some were in their first year after the Hogwarts plan was released. Additionally, some departments were excluded because they had previously not participated in assessment, either because they had not been required to, or because they had refused to. For completeness, for all analyses, we also performed between-subjects comparisons that included all departments and found identical results. For each department, we compared the timeliness of three-year assessment plans and timeliness as well as quality of annual report submissions before and after the release of the new instructions.

Materials and Procedure

The new instructions were sent to all departments at the end of the previous assessment cycle (i.e., after the last annual report had been submitted but before the new assessment plans were due) as part of the standard reminder email.

What the assessment process can uniquely offer academics is communication. Faculty do not get many opportunities to communicate their needs in a way that is compelling to administrators

---

[2]See full instructions and Hogwarts documents in an OSF repository online:
https://mfr.osf.io/render?url=https://osf.io/85muj/?action=download%26mode=render

The message did not deviate from previous reminder emails except to mention that the instructions were (a) new and (b) included example materials from Hogwarts School of Witchcraft and Wizardry. Anecdotally, several department chairs reported that reading "Hogwarts" in the email intrigued them into opening the instructions. All other interactions with the departments followed the usual protocols (e.g., responding to inquiries, offering in-person help when requested). Per standard procedure, the instructions were sent again after the assessment plans had been approved but before the first annual report was due.

The date of submission was recorded for all assessment plans and subsequent annual reports. As part of the typical assessment procedure, annual reports were rated on quality (1 = does not meet expectations, resubmission required; 2 = approaches expectations, dean's approval recommended after AAC endorses suggested changes; 3 = meets expectations, dean's approval recommended by AAC).

Results

Assessment Plans

Although the instructions were sent only a couple of weeks before the plans were due, the plans were still turned in notably earlier (M = 24.5 days late, SD = 43.0) than the previous cycle (M = 58.1 days late, SD = 83.7; see Figure 1). Although on average the plans were still overdue, it is worth noting that only 21% of the plans were on time (or early) before the new instructions and that 61% of the plans were on time (or early) after the introduction of the new instructions.
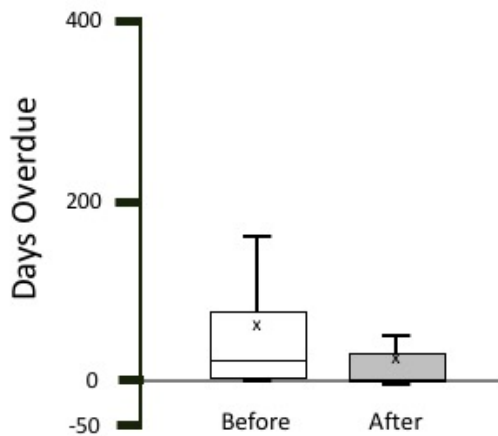
Only 21% of the plans were on time (or early) before the new instructions… 61% of the plans were on time (or early) after the introduction of the new instructions

Figure 1. Box and whisker plots of timeliness of assessment plans, new instructions significantly decreased the number of days overdue by an average of 33.6 days.

Annual Reports

By the time the first round of annual reports was due the Hogwarts documents had been circulating for almost a year. Previous participation in annual reports had been particularly low; therefore, even though every department submitted an assessment report, historical timeliness data were only available for 16 departments. For those departments, the annual reports were turned in considerably earlier (M = 9.2 days late, SD = 19.42) than the previous year (M = 77.2 days late, SD = 111.02; see Figure 2).

In addition to missing historical data, quality assessments were absent from some submitted reports because of turnover in AAC membership; therefore only nine departments' data were available for direct comparison. For these, however, the post-Hogwarts annual reports were also found to be of higher quality (M=2.89, SD=0.33) than those pre-Hogwarts (M = 2.44, SD = 0.53; see Figure 3).
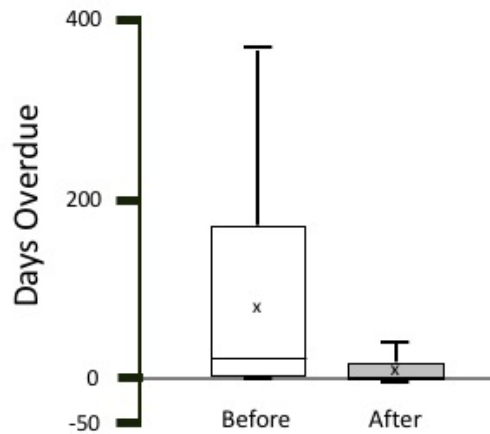
Figure 2. Box and whisker plots of timeliness of assessment reports, new instructions significantly decreased the number of days overdue by an average of 68 days.
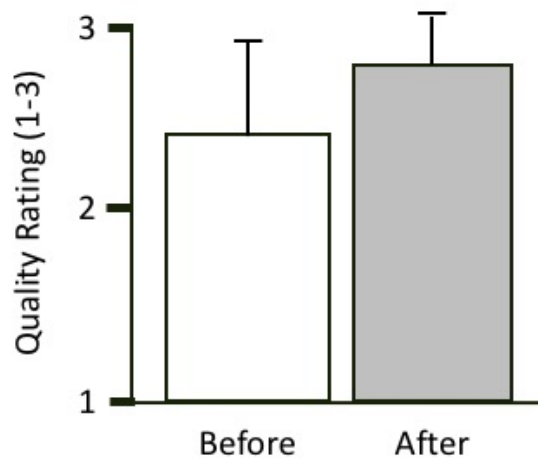
Perhaps the most important result from this endeavor is that the release of these instructions coincided with an increases in assessment participation from 22.5% of departments to 100% within a year

Figure 3. Assessment reports were rated at significantly higher quality after the release of the new instructions.

Closing the Loop

Perhaps the most important result from this endeavor is that the release of these instructions coincided with an increases in assessment participation from 22.5% of departments to 100% within a year. The assessment plans and annual reports were turned in earlier and the latter were of higher quality. Anecdotally, interactions with the departments were also smoother: Once the departments realized that we were trying to offer them a direct line of communication with the administration that could be collaborative instead of combative, they were far more inclined to participate.

It is worth noting that this endeavor was not undertaken with the usual precautions of a scientific study. While we kept everything but the instructions constant from one cycle to the next, the instructions themselves altered many things simultaneously. They referenced a popular series, they were humorous, they were written clearly and without jargon, and they stressed the importance of assessment for meeting the department's goals (rather than the administration's goals or for accreditation). Further research would need to be done to tease apart the relative

influences of each of these, but our intuition is that the Harry Potter references made them curious enough to start reading, the humor and clarity of writing made them finish reading, and the shift of purpose convinced them to follow through.

Ours is far from the first attempt at improving faculty engagement in assessment (e.g. Calegari, Sibley, & Turner, 2015; Haviland, 2009; Smith, 2005; Williams, 2013). Others have written extensively about the best ways to engender institutional change—see, for example, Calegari et al.'s discussion of Kotter's (1996) model as compared to Lewin's (1947) model—but one common finding is that increasing the desire for change on the part of the people involved is crucial. It is impossible to convince people if you cannot get them to listen, and this is fundamentally why we believe our approach was effective.

References

Banta, T. W. (Ed.). (2002). Building a scholarship of assessment. John Wiley & Sons.

Calegari, M. F., Sibley, R. E., & Turner, M. E. (2015). A roadmap for using Kotter's organizational change model to build faculty engagement in accreditation. Academy of Educational Leadership Journal, 19(3), 31.

Driscoll, A., & De Noriega, D. C. (2006). Taking ownership of accreditation: Assessment processes that promote institutional improvement and faculty engagement. Stylus Publishing, LLC..

Emil, S., & Cress, C. (2014). Faculty perspectives on programme curricular assessment: individual and institutional characteristics that influence participation engagement. Assessment & Evaluation in Higher Education, 39(5), 531–552.

Ewell, P. T. (2002). An emerging scholarship: A brief history of assessment. In T. W. Banta & Associates (Eds.), Building a scholarship of assessment (pp. 3–25). San Francisco, CA: Jossey-Bass.

Haviland, D. (2009). Leading assessment: From faculty reluctance to faculty engagement. Academic Leadership: The Online Journal, 7(1), 23.

Kotter, J. P. (1996). Leading change. Boston: Harvard Business School Press.

Lewin, K. (1947). Frontiers in group dynamics. Human Relations, 1(1), 5–41.

Rowling, J. K. (1997-2007). Harry Potter (vols. 1-7). New York, NY: Arthur A. Levine Books.

Smith, J. S. (2005). Fostering faculty engagement in departmental assessment in higher education. Assessment update, 17(4), 7–9.

Williams, J. L. (2013). Faculty engagement with learning outcomes assessment: A study of public two-year colleges in Colorado. (unpublished doctoral dissertation). University of Denver, USA.

Abstract

One of the important topics in assessment and institutional effectiveness research is the validation of assessment measures and instruments to evaluate students' learning outcomes. Using a university-wide initiative of Writing across the Curriculum as an example, this paper provides the step-by-step guidance of the mixed-methods exploratory sequential strategy in designing and validating a survey to assess one of students' learning outcomes—process writing. To generate preliminary survey items, the study started with focus groups with students, followed by content review by faculty members and writing experts. The survey was then piloted with a small sample of students, revised, and finally used with a larger sample size of students in the field. By elaborating on the steps of the mixed-methods approach in survey development, the study provides energy into the mission of accurately evaluating academic excellence and student learning in institutional research and practices.

AUTHORS

Xiaomei Song &
Michael Elftman

Central Michigan University

Survey Development in Assessing Student Learning Outcomes:
A Mixed-methods Approach

In recent years, higher education evaluation and assessment in the United States and around the world has given much attention to the use of appropriate measures in assessing student learning outcomes (SLOs) so that results based on those measures are reliable and valid (e.g., Al-Thani, Abdelmoneim, Daoud, Cherif, & Moukarzel, 2014; Ewell, 2001; NEASC, 2016; SACSCOC, 2012). As such, developing an assessment tool with reliability and validity evidence is important, particularly in university-level assessment because such assessment often involves a large number of stakeholders and aims at assessing complex, multifaceted SLOs across multiple years. Validation of instruments to evaluate SLOs has become an eminent topic in institutional research (Meyer & Zhu, 2013). Researchers in higher education have called for well-designed methodologies and approaches to examine SLOs and program outcomes (McLeod, 1992).

The purpose of this project, which included a series of studies, was to develop a reliable and valid assessment tool that would be used to investigate one of the SLOs, students' engagement in process writing. Using a university-wide initiative of Writing across the Curriculum (WAC) as an example, this paper describes the mixed-methods exploratory sequential strategy in designing and validating the survey to investigate students' writing processes. WAC is considered a high-impact educational practice (Kuh, 2008). It has been identified as the focus of Quality Enhancement Plans (QEP) in many institutions including this research context—a public, comprehensive university with more than 20,000 students within the university system of Georgia. WAC was selected for our QEP based on the data-driven needs assessment results collected from a wide population of students, faculty, staff, and alumni. Participating undergraduate programs designated a sophomore, junior, and senior-level course within their disciplines as their WAC courses. In addition to the first institutional-level SLO, "demonstrating argumentation, analysis, and synthesis skills through writing in a variety of contexts" (Georgia Southern University, 2015, p.3), the second SLO highlights students' engagement in the processes of writing including researching, drafting, revising, editing, collaborating, and reflecting (Georgia Southern University, 2015).

CORRESPONDENCE

E-mail: song5x@cmich.edu

Although process writing has been considered to be an important aspect of enhancing writing (Balgopal & Wallace, 2013; Flower, 1994), a majority of studies generally attempted to examine process writing with school-aged populations (MacArthur, Graham, & Fitzgerald, 2006). Middle school or high school students were asked to think about their processes in writing essays and they were interviewed about their writing experience (e.g., Myhill & Jones, 2007). We have been unable to identify the existence of any survey of process writing appropriate for the university setting that demonstrated reliability and validity evidence. A systematic approach to the evaluation of process writing is warranted.

As such, this project aimed to design an assessment tool with reliability and validity evidence using the mixed-methods approach. The mixed-methods approach has long been adopted in the field of measurement to construct quantitative instruments grounded in the experiences of participants (Creswell & Plano Clark, 2011). However, this approach that helps to produce rigorous tools to appropriately assess SLOs is under-researched in institutional assessment. Although the mixed-methods design has long been used to create psychometric instruments (Creswell, 2002; Hinkin, 1998), limited practical guidance is available for designing robust and rigorous surveys in an institutional context. An appropriately designed questionnaire may help to define constructs that are often multidimensional, investigate changes over time, shed light on subgroup differences, explain relationships among sets of variables, and provide in-depth information on components needing attention. Further work is needed to explicitly describe the ways of optimizing the development of instruments by using mixed-methods approaches (Onwuegbuzie, Bustamante, & Nelson, 2010). Therefore, this paper reviews the literature related to process writing and provides a step-by-step description of the mixed-methods, exploratory approach to developing the process writing survey to assess the specific SLO defined in our university initiative.

this project aimed to design an assessment tool with reliability and validity evidence using the mixed-methods approach

Students' Engagement in Process Writing

Since the 1980s, writing pedagogy has expanded from solely teaching students to concentrate on their written product to consider writing processes (Hillocks, 1986; Cumming, 1998). A review of the literature indicates that process writing evolved from a traditional, cognitive approach to a broader approach including metacognition, and to the more complex social-cognitive approach that is widely accepted today. While traditional models usually consisted of distinct steps to be followed in writing, theories in the 1980s (e.g., the Cognitive Process theory of writing by Flower & Hayes, 1980; 1981) shifted from the traditional linear sequence models to the recursive nature of writing processes. Writing was perceived to be a recursive process in which writers had the opportunity to plan, draft, edit, and revise their work. Flower and Hayes suggested four major cognitive writing processes—planning, translating, reviewing, and monitoring. Planning helps to organize ideas and brainstorm; translating takes the conceptual plan and generates texts; reviewing asks the writer to refine the text and revise content; and monitoring includes metacognitive activities of each stage and coordinates planning, translating, and reviewing. More recently, writing has been highlighted as an active, constructive, collaborative process (Flower, 1994). Flower's framework of the social-cognitive view of the writing process focuses on acts of negotiation and the insights from students' reflections, in addition to students' interpretation of tasks, feedback, and situations. Overall, current conceptualizations view writing as a constructive and contextualized process and emphasize the impact of interactions among the contextual factors on the cognitive processes of writers. The theoretical concepts of cognitive, metacognitive, and social strategies used in writing processes provide important foundations to define process writing.

Specifically, this project focused on six components of process writing for our QEP SLOs: researching, drafting, revising, editing, collaborating, and reflecting. First, the literature shows that researching, or information literacy, provides students an opportunity to reveal their understanding and interpretation of the topic by reviewing, evaluating, summarizing, or synthesizing sources (Lupton, 2004). Given the interdisciplinary nature of WAC, researching by seeking evidence can support literary interpretations, provide reasoning, present historical analyses, or identify literature gaps. Second, the literature finds that, based on the results of

researching, students begin to develop the structure of their paper by drafting, or prewriting, which includes creating concepts maps and outlines (Bahls, 2012). Drafting helps to plan out what is going to be written. Research indicates that skilled writers spend significant time organizing and planning what they are going to write (Hillocks, 1986). Students who spend little time researching and planning how to express their thoughts before writing them down may not adequately access sufficient information and ideas that could enhance the quality of their writings. The third component, revising, has a strong recursive nature. It allows students to consider their audience and continuously visit and revisit global problems of their work (e.g., argument and structure of large units of text; Bahls, 2012). This component allows writers to consider new ideas and thoughts and incorporate them into their writing. Research suggests that the process of revising helps writers to become more self-conscious and improves the quality of the final writing products (Desmet, Miller, Griffin, Balthazor, & Cummings, 2008). Fourth, editing offers opportunities to address issues in mechanics and format. Using a case study approach, Sommers (1980) found that novice writers typically revisit their writing by only fixing grammatical errors and spelling. The differences between expert and novice writers lay in the fact that experts spend significantly more time revising on global issues than novices. Fifth, as described by the social-cognitive view of the writing process (Flowers, 1994), when students draft, revise, and edit their work, students simultaneously collaborate with others. By collaborating, students review comments and feedback that they receive from others (Daniel, Gaze, & Braasch 2014; Woodrich & Fan 2017). Finally, students perform reflection, in which they consider potential changes to the draft and others' comments to develop strategies to incorporate them into a future draft (Yancey, 1998). Students reflect on feedback and determine whether the input supports their own ideas and those presented in the research they have read.

Overall, writing is a highly nonlinear activity in which writers, expert writers in particular, revisit literature and their work, often many times, over the various aspects discussed above. Although Flower's work and the social cognitive view of process writing have existed for decades, few if any questionnaires have been developed and validated to assess undergraduate students' writing process under the guidance of Flower's framework. To address this challenge, a mixed-methods, exploratory approach was used to develop a tool to investigate the SLO of process writing for institutional assessment. In addition to a step-by-step guidance, the goal of this study was to develop a preliminary set of items to capture the key components of process writing, refine the instrument, and gather preliminary validity evidence of the measure.

Validity and the mixed-methods exploratory approach

The key issue in survey design and development is to provide evidence of validity for intended uses. The contemporary perspective views validity as a multifaceted construct seeking out multiple sources of evidence, including traditionally emphasized concepts such as content, concurrent, and predictive validity (Messick, 1989; Kane, 2006). The most recent version of the AERA, APA, and NCME Standards (2014) states that an appropriate operational definition of the construct an instrument intends to measure should include a demonstration of validity evidence based on content, internal structures, response processes, and relations to other criteria. Validation is not an activity that occurs only when the survey is designed, but rather is an ongoing, dynamic process initiated at the design stage and continuing throughout development and implementation (Messick, 1989).

Thus, the mixed-methods exploratory technique has become more and more popular due to the ability to optimize survey development with validation processes (Onwuegbuzie et al., 2010). The approach consists of two sequential phases, first qualitative, then quantitative. It starts with the collection and analysis of qualitative data using various methods such as one-on-one interviews, focus groups, or direct observations to explore a phenomenon. Based on the qualitative results, researchers design and develop preliminary instrument items. The instrument's items will then be validated through a variety of forms of quantitative evidence such as reliability estimates and exploratory factor analysis. Generally speaking, the qualitative and quantitative methods are linked through the development of the survey items. In the exploratory approach, a greater emphasis is often placed on the qualitative data, which inform the development of

The key issue in survey design and development is to provide evidence of validity for intended uses

quantitative measures like Likert-form scale statements (Creswell, 2006). Creswell and Clark (2011) suggest that this approach is straightforward and easily "acceptable by quantitative-biased audiences" (p. 89).

In this project, focus group interviews with students were conducted first to generate preliminary survey items at the qualitative stage. The intention was to develop a bank of survey items that helped to identify a variety of students' activities and thinking processes representing manifestations of students' engagement in process writing. The qualitative stage also involved a group of faculty members from various disciplines to establish validity evidence in terms of its content. The faculty members conducted content analyses with the survey items in an iterative process after initial survey items were produced based on the qualitative interviewing results. Their perceptions were used to revise the instruction, content, interpretation, and wording of the survey. After that, in the second stage, the pilot study was conducted with a small sample to provide internal validity evidence of the survey. The process writing survey was finally revised for the field test with a larger sample size, and the results were reported.

The elaboration of the various steps of the mixed-methods procedure in this paper should inspire and inform other researchers to produce more solid, meaningful assessment tools

Using a mixed-methods design, this study aims to offer practical guidelines while creating a tool with validity evidence to assess process writing, which will provide a new resource for the core mission of appropriately assessing academic excellence and SLOs in an institutional context. The elaboration of the various steps of the mixed-methods procedure in this paper should inspire and inform other researchers to produce more solid, meaningful assessment tools that can be used to appropriately assess SLOs. In addition, this research helps to fill a void in the literature of the writing process by developing and validating a process writing measure.

Qualitative Stage: Focus Group

Participants

To generate preliminary survey items, 11 undergraduate students were recruited through the university public bulletin board at the library. Two focus group interviews, with one composed of three and another of eight, were conducted. The focus group participants were recruited using a purposeful sampling technique based on the QEP focus with middle- or upper-level students and their availability. Among the 11 participants, there were three sophomores, two juniors, and six seniors. Six males and five females came from a variety of academic backgrounds including general studies, political science, education, business, chemistry, nutrition and food science, nursing, English, electrical engineering, information technology, and public relations. The participants were representative of the wider undergraduate student population in terms of year of study, gender, and academic background.

Instruments

The semistructured interview protocol was designed and centered by the six components of process writing, which was identified by the QEP university committee (see Appendix A). The interview protocol essentially included two sections: the warm up question about their writing assignments (Q1) and six open-ended questions which were consistent with the six aspects of process writing (Q2-Q7). The interviews also provided the participants with opportunities to express their experiences and perceptions regarding the writing process based on their experience and reflections (Q8).

Data Collection and Analyses

The focus group interviews were conducted in January 2015. Prior to the interview, the participants were introduced to the purpose of the interview and told their participation was completely voluntary. Signed consent forms were collected after the introduction. The interviews ranged about 30–60 minutes and were voice-recorded, transcribed, and analyzed using Nvivo 9. Using the deductive coding technique (Crabtree & Miller, 1999), the initial list of themes was identified based on the six components of process writing. We then defined and modified the meaning of those themes within the process of the analysis.

Results

Through focus group interviewing with undergraduate students from different academic backgrounds, qualitative data were collected and analyzed to provide preliminary information regarding each component of process writing. Specific definitions emerged from the focus group interviews (see Table 1). The descriptions of those themes were used to form the basis of the survey constructs.

Table 1
Qualitative Themes

Theme	Descriptions
Researching	Students' behaviors to locate, read, evaluate, and cite information related to topics of writing assignments
Prewriting/drafting	Students' behaviors to organize and plan what will be written
Revising	Students' behaviors to consider audience and address global problems including clarity, reasoning, logic, audience, and flow of ideas
Editing	Students' behaviors to address mechanical, sentence structure, and format issues
Collaborating	Students' behaviors to work with other people to improve the quality of their writing assignments
Reflecting	Students' reflection on feedback and thoughts about the appropriateness and meaningfulness of feedback

Based on the interview results, previous literature, and the SLO surveys used at Georgia Southern University (e.g., ENGL 1101/1102), initial items were designed to be used and perceived as meaningful and relevant in this setting. The survey items were written in the present tense in first person narrative form to encourage participants to think about their general usage. Each item was worded in simple language to ensure the strategies were clear to participants and to avoid confusion. The survey included 21 items that examined six components of the SLO. Table 2 presents three examples of qualitative data transformed into specific survey items.

Qualitative Stage: Content Analysis

Participants

To further provide validity evidence regarding the survey content, the survey draft was presented to the QEP development committee and four experts from various disciplinary areas including nursing, journalism, English, and education. All members in the QEP development committee were invited to provide comments either collectively or individually. The development committee, 16 in total, had broad representation of faculty from each college, students, administrators, and staff (e.g., librarians). The committee members were familiar with the contextualized teaching and/or learning environment as well as the nature of the QEP outcomes.

Instrument

The survey, which included 21 items that examined the six aspects of process writing, used a 1–6 rating scale where 1 = "Never true of me" 2 = "rarely true of me (about 20% of the time)," 3 = "sometimes true of me (about 40% of the time)," 4 = "often true of me (about 60% of the time)," 5 = "usually true of me (about 80% of the time), and 6 = "always true of me." Participants were asked to evaluate the survey for redundancy, clarity, and readability of items and for suggestions or input on additional strategies. In particular, the development committee was invited to draw special attention to three aspects of the survey: 1) content—how each item measured what the QEP intended students to be able to be engaged in process writing, 2) appropriate wording to avoid misinterpretations, and 3) format and user-friendliness. This stage ensured the clarity of the survey and determined if any adjustments were required for wording or conceptual problems.

The survey items were written in the present tense in first person narrative form to encourage participants to think about their general usage. Each item was worded in simple language to ensure the strategies were clear to participants

Table 2
Creating Survey Items from Interviewing Data

Interviewing data	Theme	Survey item
"I am usually using textbook that the professors require or maybe outside source. Galileo system, with the USG system, is really helpful"	Researching	To prepare for writing, I search scholarly resources such as Galileo and narrow searches to find credible and relevant information.
"First of all, if it's about rearranging sentences, changing words, or using a better word, that is just kind of 'I agree' 'I will go with that'… so a lot of times, if I get something where a person's question is about content itself, then there is a lot more of kind of reflection on that piece, and trying to clarify. How can I get my point across, but understand that they are not getting it right now? How can I make it clear?"	Reflecting	I read and review the comments that I received to see if they sound right or if they make sense to me.
"I come from a family of teachers. My mom usually helps me with ideas. Get feedback from them. They are really supportive of it."	Collaborating	I seek feedback and comments from instructors, peers [e.g., classmates, lab mates, roommates, friends or family members on my draft(s)].

Survey items were revised several times in an iterative process based on input and feedback from the QEP committees and other experts

Results

　　Survey items were revised several times in an iterative process based on input and feedback from the QEP committees and other experts. This process was intended to produce preliminary survey items to be used at the institutional level for the pilot study. Based on the comments, 21 items were developed. Special care was taken to ensure that each item only reflected one type of question to avoid the use of double-barreled items. Table 3 presents examples of the revised items.

Table 3
Revising Survey Items

Original item	Concern	Revised item
To prepare for writing, I highlight important sections, bullet point the key issues, take notes in my reading, or write annotated bibliography.	Annotated bibliography can be one type of writing assignment.	To prepare for writing, I highlight important sections, bullet point the key issues, or take notes in my reading.
Before I start writing, I think about the key points and visualize a concept map.	Students may not understand the meaning.	Before I start writing, I think about the key points and visualize their order or relationships.
I read my writing carefully to make sure there are no errors in citations.	Changing to a strong verb	I read my writing carefully to eliminate errors in citations.
I go to the library, the Writing Center, or Academic Success Center to get help on my writing.	Changing from a general to specific word	I go to the library, the Writing Center, or Academic Success Center to get guidance on my writing.

Quantitative Stage: Pilot Study

Participants

After survey items were created and revised, the survey was piloted with eight intact classes, a total of 277 undergraduate students with various academic backgrounds including engineering, biology, psychology, international trade, and English. Four responses with missing values exceeding 10% of the total number of 21 process writing items were removed from the database. Among the 273 valid responses, 82 came from College of Engineering and Information Technology, 64 College of Liberal Arts and Social Sciences, 44 from College of Business Administration, 5 from College of Education, 2 from College of Public Health, 36 from College of Science and Mathematics, and 37 from College of Health and Human Sciences. Three students did not report their majors. One hundred forty-five were males, 126 were females, and two did not indicate their genders. In addition, there were 24 freshmen, 117 sophomores, 73 juniors, and 59 seniors.

Data Collection and Analysis

Instructors who taught in this university were contacted through personal connections. They were invited to help with data collection, either by distributing the survey in their class or allowing the researcher to go to their class to distribute the survey. Students were informed that under no circumstances would their answers for the survey be released to anyone else but the researcher. Students were told that their frame of reference for responding to the survey statements should be their engagement in process writing in general, rather than any particular type of assignment.

In order to provide the reliability and validity evidence of the process writing survey, a series of data analyses were computed, including descriptive statistics, reliability estimates, and exploratory factor analysis. Item-level descriptive statistics were calculated first to determine how the participants in the pilot study reported their engagement in process writing. Normality of the survey was examined because it is an early step in factor analysis. A nonnormal distribution of the items could degrade the correlations among variables and consequently weaken factor analysis. Reliability estimates were also calculated to investigate the homogeneity of subscales because it is a prerequisite for subsequent factor analysis. As exploratory factor analysis is a common way in the early stages of scale development for data reduction (Kelloway, 1995), it was used to determine the items that load best on each factor. The sample size met the minimum recommended requirement, item ratio of 10 to 1, for obtaining stable factor solutions (Burton & Mazerolle, 2011).

The reliability estimates of most of the subscales were above .60, except the subscale of researching (α = .48), which is somewhat low

Results

The means of the 21 items ranged from 2.5 to 4.92 and the standard deviations from 1.05 to 1.52. All skewness and kurtosis values ranged between +1 and -1, except one (Researching 3), and all were within the accepted limits (+ 2), indicating that the responses for individual items seemed to be normally distributed. Therefore, all the survey items were kept for the subsequent analyses. Cronbach's alpha was computed to discover the level of internal consistency for the subscales of the survey. The reliability estimates of most of the subscales were above .60, except the subscale of researching (α = .48), which is somewhat low.

Table 4
Descriptive Statistics at the Components Level

	# of Items	N	M	SD	Skewness	Kurtosis	Reliability
Researching	4	272	4.34	0.81	-0.26	0.04	0.48
Drafting	3	273	4.29	1.10	-0.29	-0.35	0.78
Revising	4	273	3.59	1.10	0.11	-0.35	0.71
Editing	3	273	4.48	1.03	-0.52	0.13	0.63
Collaborating	4	273	3.48	1.21	-0.19	-0.52	0.84
Reflecting	3	273	4.72	1.03	-0.77	0.40	0.87

To understand how the 21 items clustered with their respective subscales, exploratory factor analysis was performed. The measure of Kaiser-Meyer-Olkin (KMO) and Bartlett's Test was .87, indicating the adequacy of the sample size. After that, four major issues were considered: (a) factor model (common factor analysis vs. components analysis), (b) rotation (orthogonal vs. oblique), (c) number of factors, and (d) interpretation (see the comprehensive review paper by Fabrigar, Wegener, MacCallum, & Strahan, 1999). In this pilot study, Maximum Likelihood and a Varimax solution were used because they maximized interpretations (see a review by Fabrigar et al., 1999). An examination of initial eigenvalues indicated that six factors had eigenvalues greater than 1.0. The scree plot also found the curve decreases and straightened out at the 6th point. Those six factors explained 67.9% of the variances. Considering the low reliability estimate of the subscale of researching, the four items under researching were removed each time because these items were unlikely to provide clear, meaningful information to explain the latent variables under investigation. The results after removing each of the research items were compared, but they still did not provide clearer patters. The rotated factor matrix was reported in Table 5, which included all the 21 items with loadings above .35. The goal in these analyses were to generate a set of discriminating items to be included in the final instrument.

13 of the 21 items loaded on the expected factors. Four Collaborating items loaded on factor 1, three Editing items on factor 2, three Drafting items on factor 3, and three Reflecting items on factor 4

Table 5
Exploratory Factor Analysis

	Factor					
	1	2	3	4	5	6
Researching1			0.39		0.36	
Researching2					0.64	
Researching3						0.73
Researching4			0.47			
Drafting1			0.71			
Drafting2			0.74			
Drafting3			0.67			
Revising1	0.36				0.41	
Revising2		0.45			0.43	
Revising3		0.71				
Revising4		0.46				
Editing1		0.61				
Editing2		0.81				
Editing3		0.38				
Collaborating1	0.70					
Collaborating2	0.81					
Collaborating3	0.84					
Collaborating4	0.55					
Reflecting1				0.65		
Reflecting2				0.84		
Reflecting3				0.81		

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.

[a]Rotation converged in 6 iterations.

As shown in Table 5, 13 of the 21 items loaded on the expected factors. Four Collaborating items loaded on factor 1, three Editing items on factor 2, three Drafting items on factor 3, and three Reflecting items on factor 4. However, four Researching items loaded on three factors and

RESEARCH & PRACTICE IN ASSESSMENT

three Revising items clustered together with Editing. Although the results provided sufficient evidence for internal consistency of the components and dimensionality of the survey, the items under Researching and Revising did not load the components as expected. As such, the items related to Researching and Revising were revisited while the other 13 items were retained in the process writing survey.

The items related to Researching and Revising were brought to a wide audience including faculty members (engineering, nursing, history, and journalism) and department chairs (communication arts and history). The four items regarding researching did not seem to measure Researching in a reliable and valid manner. Although researching was defined as "seeking evidence" (Lupton, 2004), the definition of "evidence" seems vary dramatically between different disciplines. While many papers contain descriptions of assignments by including a research component in graduate programs, no one has proposed a coherent definition for research that can be used in the context of undergraduate writing assignments. The conversations with faculty members from various disciplinary backgrounds found different definitions, experiences, and expectations toward researching. For example, the engineering faculty member found students mostly used Google while he expected students to use scholarly resources including Google Scholar for researching. Conversely, the faculty in journalism recognized the importance of using Google search to find background information related to human sources. Although the history faculty member recommended secondary resources such as journal articles, he generally found it impractical to use human sources for researching. For some areas, researching goes beyond information literacy. Researching indicates not only seeking empirical evidence, but also demonstrating data collection, analysis, and synthesis skills for theory or practical implications.

The four items that intended to measure Revising were also problematic. One possible reason may be due to the situation that students did not distinguish revising from editing during the process of their writing. They may check errors or inconsistencies only at the local level such as word choice, a missed word, or sentence structures. Students do not revise the paper as a whole, considering the strengths and weaknesses of arguments. Based on those comments, the items related to Researching and Revising were revised (see Table 6).

Quantitative Stage: Field Study

Participants

The survey (see Appendix B) was distributed in 36 classes. Eight hundred forty-one students submitted the survey and there were 785 valid responses. Among the valid responses, 256 were juniors (32.6%) and 523 were seniors (66.6%); 346 were males (44.1%) and 430 were females (54.8%). Students came from a variety of programs and colleges including the College of Engineering and Informational Technology (12.6%), Collage of Health and Human Sciences (20.4%), College of Education (5.9%), College of Business Administration (21%), College of Science and Mathematics (13%), College of Public Health (1.3%), and College of Liberal Arts and Social Sciences (25%). Seven participants did not specify their college affiliation.

Data Collection and Analysis

Through the university system, instructors were invited to distribute the survey in their classes. Written instructions were provided to instructors to explain the purpose and structure of the survey and the approach to return the responded surveys. After that, a series of data analyses were computed, including descriptive statistics, reliability estimates, and exploratory factor analysis.

Results

All skewness and kurtosis values of the 21 items were within the acceptable limits (+ 2). Therefore, all the survey items were kept for the subsequent analyses. The process writing survey showed sufficient internal consistency evidence for the overall survey (α = .91) and its six components (ranging from a low of .70 to a high of .85).

The items related to Researching and Revising were brought to a wide audience including faculty members (engineering, nursing, history, and journalism) and department chairs (communication arts and history)

Table 6
Refining Survey Items

Revised survey item	Refined survey item
Researching 1: To prepare for writing, I try to collect as much relevant information as possible.	Researching 1: To prepare for writing, I use credible resources (e.g., GALILEO, Google Scholar, or books) to develop the topic or support the argument.
Researching 2: To prepare for writing, I search scholarly resources such as GALILEO and narrow searches to find credible and relevant information.	Researching 2: I read sources critically to see whether they are based on opinions, facts, or empirical evidence.
Researching 3: To prepare for writing, I search resources such as Google for information.	Researching 3: To prepare for writing, I keep track of the information of sources so I can cite them properly.
Researching 4: To prepare for writing, I highlight important information, bullet point key issues, or take notes in my reading.	
Revising 1: I write multiple versions for my assignment instead of finishing my paper in one sitting.	Revising 1: I write multiple versions for my assignment instead of finishing my paper in one sitting.
Revising 2: I reorganize what I write by moving around ideas, sentences, and/or paragraphs to make it more logical.	Revising 2: I reorganize what I write by moving around ideas, sentences, and/or paragraphs to make it more logical.
Revising 3: I check to see if sentences make sense together, add sentences to create better flow or connection, and/or make links between different parts of writing.	Revising 3: I write more than one draft in order to improve the overall structure of my writing assignment.
Revising 4: When writing I think about my readers and adjust the way I describe things or expressions.	Revising 4: I write more than one draft to clarify the points/ideas that I discuss in my writing assignment.
	Revising 5: I consider the audience of my writing assignment and adjust the way that I write.

It also filled a void in the literature by developing a set of Likert-scale questions to measure Flower's social-cognitive view of the writing process

In the field study, Principal Axis Factoring and a Varimax solution were used. An examination of initial eigenvalues indicated that six factors had eigenvalues greater than 1.0. The scree plot also found the curve decreases and straightened out at the 6th point. Those factors explained 70.3% of the variances. Rotate factor matrix was reported in Table 7. Revising was the component that accounted for the largest variance of process writing. The process writing survey showed validity evidence by examining the internal structure of the survey through Exploratory Factor analysis, except one item Audience Consideration under Revising—"I consider the audience of my writing assignment and adjust the way that I write." This item will need closer monitoring to see how the loading may change in the future.

Discussion

This research demonstrated how a mixed-methods approach helped to develop an assessment tool with reliability estimates and validity evidence so that students' engagement in the process writing could be appropriately investigated. It also filled a void in the literature by developing a set of Likert-scale questions to measure Flower's social-cognitive view of the writing process. Following the established mixed-methods survey development steps, four separate steps of research were used. The Qualitative Stage: Focus Group involved item generation and development. The Qualitative Stage: Content Analysis was to provide evidence

Table 7
Exploratory Factor Analysis

	Factor					
	1	2	3	4	5	6
Researching1						0.67
Researching2						0.63
Researching3						0.48
Drafting1					0.62	
Drafting2					0.78	
Drafting3					0.69	
Revising1	0.65					
Revising2	0.40					
Revising3	0.92					
Revising4	0.88					
Revising5	0.20	0.29	0.33	0.05	0.25	0.14
Editing1			0.82			
Editing2			0.85			
Editing3			0.52			
Collaborating1				0.65		
Collaborating2				0.75		
Collaborating3				0.76		
Collaborating4				0.41		
Reflecting1		0.66				
Reflecting2		0.83				
Reflecting3		0.70				

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.

[a]Rotation converged in 7 iterations.

related to content validity of survey items. The development and refinement of a content valid instrument was achieved by a rational analysis of the instrument by content experts who were familiar with the research context and subject. Those experts examined the questionnaire items for readability, clarity, and comprehensiveness and came to agreement about which items should be included in the questionnaire. This step enhanced collaboration, engagement, and buy-in among faculty members. The Quantitative Stage: Pilot Study provided information regarding internal structure. The construct of interest process writing had many dimensions which formed different domains of a general attribute, therefore, factor analysis was employed. In the analysis of internal structure, items that measured a particular dimension within a construct of interest were more highly related to one another than to those measuring other dimensions. While the pilot study tested the internal structure of 21 items, the Qualitative Stage: Field Study confirmed the six dimensions of the survey for all but one item, Audience Consideration. Overall, the survey demonstrated evidence in terms of content validity (steps one and two), internal structure (steps three and four), and reliabilities above the recommended level for new scales (steps three and four). Therefore, the study established empirical evidence of the appropriateness of the process writing survey guided by Flower's theory to be used in the current context.

Given the high relevance of survey development to this community, this work has important methodological implications and informs practical applications to address challenging questions in institutional assessment. First, the paper provides explicit information and step-by-

The survey demonstrated evidence in terms of content validity (steps one and two), internal structure (steps three and four), and reliabilities above the recommended level for new scales (steps three and four)

step guidance about enhancing the reliability and validity evidence of a survey assessing student learning. The information goes beyond a specific setting in designing a particular survey and provides practical, methodological guidance about the application of fundamental concepts in assessment (e.g., reliability and validity) in designing indirect measures and investigating SLOs. The mixed-methods exploratory approach can be a useful tool to investigate multidimensional constructs for institutional assessment. The qualitative stage is to conceptualize the construct of interest based on a thorough review of the literature and information provided by key informants and experts. The information is used to identify and describe behaviors that underlie the construct and develop initial instrument items. After that, the quantitative stage pilots the initial items and provides construct-related validity evidence. It helps to revise initial items through field testing those items. Second, the process writing survey may be used by other schools and institutions as a basis to be validated in their contexts and evaluate how students are engaged in writing processes when they write writing assignments. The survey can generate ideas for other contexts and help organizations to uncover potential problems and identify specific areas that need special attention. Finally, the process writing survey will allow for greater use of Flower's social-cognitive theory of process writing. Although the model has shown promise and been cited considerably in previous literature, the lack of direct measures of process writing from this model has limited its full application.

This current series of studies addresses a gap in process writing literature by describing how to design and develop a rigorous, meaningful scale to evaluate SLOs

Conclusion

Overall, the results provide information concerning internal consistency and construct validity of the components of process writing. The results show that process writing involves the specifics of the actual writing process encompassing cognitive, metacognitive, and social strategies. A single survey item, which is often used to examine student process writing, underestimates the complexity of process writing, and hence, does not provide an accurate estimate of the SLO. The mixed-methods approach in survey development, a relatively underutilized method, helps to produce tools with reliability and validity evidence that can be used to appropriately assess SLOs in institutional assessment. This current series of studies addresses a gap in process writing literature by describing how to design and develop a rigorous, meaningful scale to evaluate SLOs.

Although the project provided step-by-step guidance on how to use the mixed-methods approach to develop a survey, there are several limitations and more research is needed to further refine and validate the process writing survey using more complicated methodologies. The possible limitations of the sampling strategy and use of focus groups should be noted. Future research may test the revised survey using confirmatory factor analysis. Different from exploratory factor analysis that was used in the quantitative stage in this project, confirmatory factor analysis tests a theory (e.g., Flower's social cognitive theory) between overserved variables (survey items) and their underlying, latent variables. Future studies also need to demonstrate convergent validity and divergent validity. Such validity evidence can be established based on a predictable pattern of relationships with other variables such as students' writing products or attitudes. Research may also consider using Item Response Theory to measure the relevance and difficulty of survey item content, evaluate appropriateness of response categories, and examine item redundancy. In addition, Onwuegbuzie and his collaborators (2010) strongly recommend including open-ended, qualitatively-based items with quantitative instruments. They affirm the importance of involving "a comprehensive evaluation both of the product and the process" (p. 67).

On the whole, the process writing survey that was constructed within the current context needs to be validated if used in other contexts. Future research also needs to examine the nature of audience consideration. Another concern of this project revolved around the question whether those mental and behavioral processes can be validly elicited by merely a self-reported questionnaire. Multiple data collection methods, such as using retrospective think-aloud protocol, are recommended in investigating the use of process writing in specific contexts.

Appendix A: Structured Interview Protocol

1. Warm-up question: Can you introduce yourself? Can you tell me the recent writing assignment you have done or you are still working on? It might be a lab report, essay, summary, or short-answer question.
2. In the following, we are going to reflect on the steps or actions that you took in your writing assignment. Before you began to write, how did you do research and get yourself prepared?
3. How did you draft? How did you spend time making an outline or drafting the structure?
4. How did you revise your writing assignment? How did you reorder, delete or add new material? If so, how much and why?
5. How did you edit your writing? Before you submitted your writing, how did you go through the whole paper to proofread and check grammar and spelling?
6. What social supports did you use for writing? How did you discuss ideas with others such as instructors, librarians, or tutors at the writing center?
7. How did you reflect on your writing? How did you reflect on responses which were given after peer review?
8. Do you have any additional comments that you would like to share with us?

Appendix B: Process Writing Survey

1. To prepare for writing, I use credible resources (e.g., GALILEO, Google Scholar, or books) to develop the topic or support the argument.
2. I read sources critically to see whether they are based on opinions, facts, or empirical evidence.
3. To prepare for writing, I keep track of the information of sources so I can cite them properly.
4. Before I start writing, I think about key points and visualize their order/relationships.
5. I spend time brainstorming and creating a web of ideas in my mind.
6. I develop and group ideas, list supporting arguments, and/or identify pros and cons.
7. I write multiple versions for my assignment instead of finishing my paper in one sitting.
8. I reorganize what I write by moving around ideas, sentences, and/or paragraphs to make it more logical.
9. I write more than one draft in order to improve the overall structure of my writing assignment.
10. I write more than one draft to clarify the points/ideas that I discuss in my writing assignment.
11. I consider the audience of my writing assignment and adjust the way that I write.
12. Before I turn in my writing, I go through my whole paper to check my word usage, grammar, punctuation, and spelling.
13. I edit my writing assignment carefully to ensure proper word choice.
14. I edit my writing assignment by following the disciplinary style guide (e.g., APA, MLA, Chicago, or specific disciplinary writing format).
15. I seek feedback and comments from instructors, peers (e.g., classmates, lab mates, roommates, and friends), or family members on my draft(s).
16. I talk with instructors, peers, or family members about my thoughts on my assignment to get outsiders' views.
17. When writing an assignment, I consult my instructors, peers, or family members for assistance or directions.
18. I go to the library, the Writing Center, or Academic Success Center for guidance on my writing.
19. I reflect on the feedback I received from instructors, peers (e.g., classmates, lab mates, roommates, and friends), and/or family members on my draft(s).
20. I read and review the comments that I received to see if they sound right or if they make sense to me.
21. I try to figure out whether the comments I received fit in the flow of my paper/writing or into the big picture in my writing.

References

Al-Thani, S. J., Abdelmoneim, A., Daoud, K., Cherif, A., & Moukarzel, D. (2014). Assessment of student learning outcomes for assurance of learning at Qatar University. Journal of Assessment and Institutional Effectiveness, 4(2), 116–136.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association.

Bahls, P. (2012). Student writing in the quantitative disciplines: A guide for college faculty. San Francisco, CA: Jossey-Bass.

Balgopal, M., & Wallace, A. (2013) Writing-to-learn, writing-to-communicate, and scientific literacy. The American Biology Teacher, 75(3), 170–175.

Burton, L. J., & Mazerolle, S. M. (2011). Survey instrument validity Part I: Principles of survey instrument development and validation in Athletic Training education research. Athletic Training Education Journal, 6(1), 27–35.

Crabtree, B., & Miller, W. F. (1999). A template approach to text analysis: Developing and using codebooks. In B. Crabtree & L. Willam. (Eds.), Doing qualitative research (pp. 163–177). Newbury Park, CA: Sage.

Creswell, J. W. (2002). Educational research: Planning, conducting, and evaluating quantitative and qualitative Research. Saddle River, NJ: Prentice Hall.

Creswell, J. W., & Plano Clark, V. L. (2011). Designing and conducting mixed methods research. Los Angeles: Sage.

Cumming, A. (1998). Theoretical perspectives on writing. Annual Review of Applied Linguistics, 18, 61–78.

Daniel, F., Gaze, C. M., & Braasch, J. L. G. (2014). Writing cover letters that address instructor feedback improves final papers in a research methods course. Teaching of Psychology, 42(1), 1–5.

Desmet, C., Miller, D. C., Griffin, J., Balthazor, R., & Cummings, R. (2008). Reflection, revision, and assessment in First-year Composition ePortfolios. The Journal of General Education, 57(1), 15–30.

Ewell, P. (2001). Accreditation and Student Learning Outcomes: A proposed point of Departure. CHEA Occasional Paper.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of Exploratory Factor Analysis in psychological research. Psychological Methods, 4(3), 272–299.

Flower, L. (1994). The construction of negotiated meaning: A social cognitive theory of writing. Carbondale: Southern Illinois University Press.

Flower, L., & Hayes, J.R. (1980). The cognition of discovery: Defining a rhetorical problem. College Composition and Communication, 31, 21–32.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. College Composition and Communication, 32, 365–87.

Georgia Southern University. (2015). Georgia Southern Eagles! Write! Write! Write!. Unpublished internal document.

Hillocks, G. (1986). Research on written composition. Urbana, IL: National Council of Teachers of English.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. Organizational Research Methods, 1, 104–121.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp.17–64). Westport, CT: American Council on Education/Praeger.

Kelloway, E. K. (1995). Structural Equation Modeling in perspective. Journal of Organization Behaviors, 16, 215–224.

Kuh, G. D. (2008) High-impact educational practices: What they are, who has access to them, and why they matter. Washington DC: AAC&U. Retrieved September 19, 2017 from http://provost.tufts.edu/celt/files/High-Impact-Ed-Practices1.pdf

Lupton, M. (2004). The learning connection: Information literacy and the student experience. Adelaide, SA: Auslib Press.

MacArthur, C. A., Graham, S., & Fitzgerald. J. (2006). Handbook of Writing Research. New York: The Guilford Press.

McLeod, P. J. (1992). So few medical schools, so many clerk rating systems! CMAJ: Canadian Medical Association Journal, 146, 2161–2164.

Messick, S. (1995). Validity of psychological assessment. American Psychologist, 50, 741–749.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational Measurement (3rd ed., pp.13–103). New York: American Council on Education/Macmillan.

Meyer, P., & Zhu, S. (2003). Fair and equitable measurement of student learning in MOOCs: An introduction to Item Response Theory, scale linking, and score equating. Research & Practice in Assessment, 8, 26–39.

Myhill, D., & Jones. S. (2007). More than just error correction: Students' perspectives on their revision processes during writing. Written Communication, 24(4), 323–43.

New England Commission of Higher Education. (2016). Standards for Accreditation. Retrieved September 19, 2017 from https://cihe.neasc.org/standards-policies/standards-accreditation/standards-effective-july-

Onwuegbuzie, A. J., Bustamante, R. M., & Nelson, J. A. (2010). Mixed research as a tool for developing quantitative instruments. Journal of Mixed Methods Research, 4, 56¬–78.

Southern Association of Colleges and Schools Commission on Colleges. (2012). Resource Manual for Principles of Accreditation. Retrieved September 15, 2018 from http://www.sacscoc.org/pdf/Resource%20Manual.pdf

Sommers, N. (1980). Revision strategies of student writers and experienced adult writers. College Composition and Communication, 31, 378–388.

Woodrich, M., & Fan, Y. (2017). Google Docs as a tool for collaborative writing in the Middle School classroom. Journal of Information Technology Education: Research, 16, 391–410.

Yancey, K. B. (1998). Reflection in the writing classroom, Logan: Utah State University Press.