RESEARCH & PRACTICE IN ASSESSMENT

VOLUME FIFTEEN | WINTER 2020 | RPAjournal.com | ISSN # 2161-4120

RPA

A PUBLICATION OF THE VIRGINIA ASSESSMENT GROUP

RESEARCH & PRACTICE IN ASSESSMENT

CALL FOR PAPERS

Manuscripts submitted to RPA may be related to various higher education assessment themes, and should adopt either an assessment measurement or an assessment policy/foundations framework. Contributions are accepted at any time and will receive consideration for publishing. Manuscripts must comply with the RPA Submission Guidelines and be submitted to our online manuscript submission system found at rpajournal.com/authors/.

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 published twice annually. Research & Practice in Assessment is listed in Cabell's Directory and indexed by EBSCO, ERIC, Gale, and ProQuest.

History of Research & Practice in Assessment

Research & Practice in Assessment (RPA) evolved over the course of several years. Prior to 2006, the Virginia Assessment Group produced a periodic organizational newsletter. The purpose of the newsletter was to keep the membership informed regarding events sponsored by the organization, as well as changes in state policy associated with higher education assessment. The Newsletter Editor, a position elected by the Virginia Assessment Group membership, oversaw this publication. In 2005, it was proposed by the Newsletter Editor, Robin Anderson, Psy.D. (then Director of Institutional Research and Effectiveness at Blue Ridge Community College) that it be expanded to include scholarly articles submitted by Virginia Assessment Group members. The articles would focus on both practice and research associated with the assessment of student learning. As part of the proposal, Ms. Anderson suggested that the new publication take the form of an online journal.

The Board approved the proposal and sent the motion to the full membership for a vote. The membership overwhelmingly approved the journal concept. Consequently, the Newsletter Editor position was removed from the organization's by-laws and a Journal Editor position was added in its place. Additional by-law and constitutional changes needed to support the establishment of the Journal were subsequently crafted and approved by the Virginia Assessment Group membership. As part of the 2005 Virginia Assessment Group annual meeting proceedings, the Board solicited names for the new journal publication. Ultimately, the name Research & Practice in Assessment was selected. Also as part of the 2005 annual meeting, the Virginia Assessment Group Board solicited nominations for members of the first RPA Board of Editors. From the nominees Keston H. Fulcher, Ph.D. (then Director of Assessment and Evaluation at Christopher Newport University), Dennis R. Ridley, Ph.D. (then Director of Institutional Research and Planning at Virginia Wesleyan College) and Rufus Carter (then Coordinator of Institutional Assessment at Marymount University) were selected to make up the first Board of Editors. Several members of the Board also contributed articles to the first edition, which was published in March of 2006.

After the launch of the first issue, Ms. Anderson stepped down as Journal Editor to assume other duties within the organization. Subsequently, Mr. Fulcher was nominated to serve as Journal Editor, serving from 2007-2010. With a newly configured Board of Editors, Mr. Fulcher invested considerable time in the solicitation of articles from an increasingly wider circle of authors and added the position of co-editor to the Board of Editors, filled by Allen DuPont, Ph.D. (then Director of Assessment, Division of Undergraduate Affairs at North Carolina State University). Mr. Fulcher oversaw the production and publication of the next four issues and remained Editor until he assumed the presidency of the Virginia Assessment Group in 2010. It was at this time Mr. Fulcher nominated Joshua T. Brown (Director of Research and Assessment, Student Affairs at Liberty University) to serve as the Journal's third Editor and he was elected to that position.

Under Mr. Brown's leadership Research & Practice in Assessment experienced significant developments. Specifically, the Editorial and Review Boards were expanded and the members' roles were refined; Ruminate and Book Review sections were added to each issue; RPA Archives were indexed in EBSCO, Gale, ProQuest and Google Scholar; a new RPA website was designed and launched; and RPA gained a presence on social media. Mr. Brown held the position of Editor until November 2014 when Katie Busby, Ph.D. (then Assistant Provost of Assessment and Institutional Research at Tulane University) assumed the role after having served as Associate Editor from 2010-2013 and Editor-elect from 2013-2014.

Ms. Katie Busby served as RPA Editor from November 2014-January 2019 and focused her attention on the growth and sustainability of the journal. During this time period, RPA explored and established collaborative relationships with other assessment organizations and conferences. RPA readership and the number of scholarly submissions increased and an online submission platform and management system was implemented for authors and reviewers. In November 2016, Research & Practice in Assessment celebrated its tenth anniversary with a special issue. Ms. Busby launched a national call for editors in fall 2018, and in January 2019 Nicholas Curtis (Director of Assessment, Marquette University) was nominated and elected to serve as RPA's fifth editor.

Published by:
VIRGINIA ASSESSMENT GROUP | virginiaassessment.org
Publication Design by Patrice Brown |

TABLE OF CONTENTS

Expanding Our Boundaries in Assessment

"*We cannot become what we want to be by remaining what we are.*" - Max DePree

"As we conclude 2020, we find ourselves looking into the new year with a sense of hope. We hope that the pandemic will abate with the distribution of new vaccines. We hope that we can rebuild much of what we have lost. We hope that we can return to some sense of normalcy. Yet, as we kindle this hope, we would also be wise to consider how we might use this time to continue expanding our practices outside of what has been "traditional" in our work. While we might indeed be able to return to the same pre-pandemic practices in higher education, that doesn't mean we must or that we should. Let us all take time to consider what we might do to improve our work and what we might add to our toolkit as we step back into our physical world. This issue of RPA includes seven peer reviewed articles that address a variety issues in our field that might encourage us to expand our boundaries.

Smith and Finney provide compelling rationale for considering the impact of program theory and implementation fidelity in our work; providing a meaningful example through an ethical reasoning program. Chase shares her work exploring how students can be more involved in classroom-level assessment and the numerous benefits that might accrue from such a practice. Jones and Phillips provide a lens for considering how traditional theories underlying many program review practices may undermine diversity, equity, and inclusion efforts, particularly at HBCUs. Clark, Luo, and Smith examine the effectiveness of and provide recommendations for embedding a dedicated writing assessor while allowing faculty to focus on disciplinary content. Singer-Freeman and Robinson detail their work to explore grand challenges in higher education assessment and provide some insight into their ongoing work on how to begin addressing those challenges. Davis, Biddle, and Hall share their work to examine faculty member experiences with internal processes within the program review process. Finally, Tucker, Drummound, and Ostrogorsky provide an account of their experience developing and working with a rubric for program-level assessment plans and reports.

I hope this issue of *Research & Practice in Assessment* provides you with some inspiration as you seek to expand your professional boundaries in the new year.

Regards,

Nicholas Curtis

Editor-in-Chief,
Research & Practice in Assessment

Abstract

Higher education institutions struggle to demonstrate learning improvement (Banta, Jones, & Black, 2009; Banta & Blaich, 2011; Jankowski, Timmer, Kinzie, & Kuh, 2018). We showcase how student learning outcomes assessment processes can benefit from strong program theory and implementation fidelity data. In our example, faculty articulated the etiology of the distal outcome of acting ethically, which allowed for specification and measurement of the intermediate student learning outcomes. Faculty specified research-informed curriculum and pedagogy to influence the intermediate outcomes and ultimately the distal outcome. By articulating the program theory, faculty were able to assess both the intermediate outcomes for gains and their associated curriculum for implementation fidelity. Faculty could then identify what aspects of programming required changes to evidence learning improvement. Thus, we argue that program theory and implementation fidelity should be prominent components of higher education outcomes assessment processes to address the dearth of empirically supported learning improvement.

LEARNING IMPROVEMENT

AUTHORS

Kristen L. Smith, Ph.D.
Macmillan Learning

Sara J. Finney, Ph.D.
James Madison University

Elevating Program Theory and Implementation Fidelity in Higher Education: Modeling the Process via an Ethical Reasoning Curriculum

Improving student learning in higher education is challenging. Few universities or colleges have used outcomes assessment data to demonstrate learning improvement (Banta, et al., 2009; Banta & Blaich, 2011; Jankowski, et al., 2018; Kushimoto, 2010). Hence, assessment practitioners are critically reflecting on their practices and developing strategies to address this shortcoming (Coates, 2016; Fulcher & Prendergast, 2019; Fulcher et al., 2017; Mathers, et al., 2018; Smith et al., 2018). Expanding traditional outcomes assessment practices could increase the likelihood of positively impacting student learning. More specifically, by articulating strong program theory (Pope, et al. 2019) and collecting implementation fidelity data (Smith et al., 2019), faculty and student affairs practitioners should be able to identify what aspects of programming (i.e., educational interventions) require changes to achieve learning improvement.

In this paper, we describe how to incorporate strong program theory and implementation fidelity into assessment practice via five steps. To illustrate these steps, we provide an example of an ethical reasoning program at our institution. The processes we describe can be applied to any academic (e.g., Meixner et al., 2020) or student affairs (e.g., Fisher et al., 2014; Gerstner & Finney, 2013) educational program.

CORRESPONDENCE

Email
kristen.smith@macmillan.com

Expanding Higher Education Assessment Practice to Include Strong Program Theory

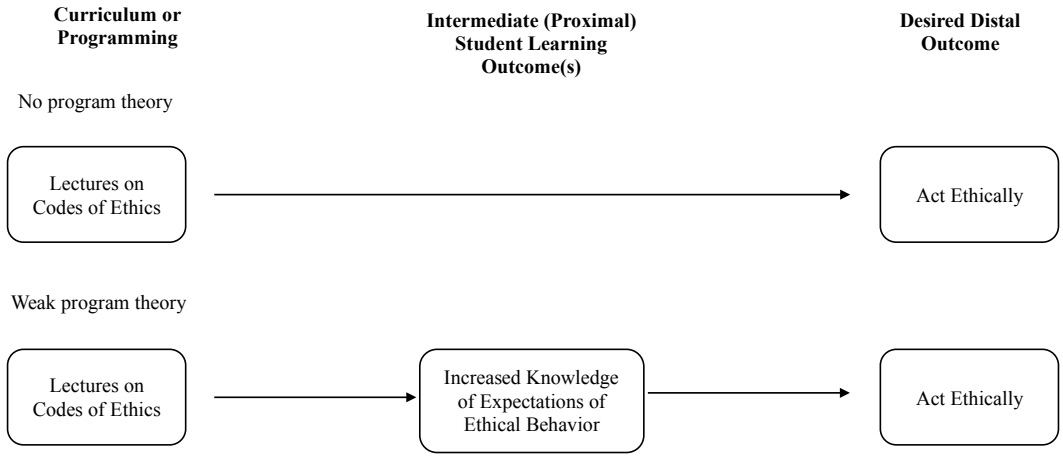
As faculty or student affairs practitioners, we are expected to design effective educational programs and assess their effectiveness (Coates, 2016; Finney & Horst, 2019a, 2019b; Leathwood & Phillips, 2000; U.S. Department of Education, 2006). However, the emphasis on gathering and reporting assessment data can distract from the equally important responsibility of designing intentional programming (e.g., curricula, activities) informed by theory and research. Creating such evidence-informed programming can be a daunting task, especially given the lack of practical guidance (Pope, et al., 2019). Yet, the use of theory and research to articulate strong program theory can inspire high-quality, valuable assessment practice.

The use of theory and research to articulate strong program theory can inspire high-quality, valuable assessment practice.

Program theory is defined as "the construction of a plausible and sensible model of how a program is supposed to work" (Bickman, 1987, p. 5). Program theory "clarifies the set of cause-and-effect relationships" believed to connect the things students do (i.e., programming) to the outcomes they are expected to achieve (Bickman, 1987, p. 5). Strong program theory is evidence-based and articulates coherent links between curriculum and/or pedagogies and student learning outcomes (SLOs). Weak program theory is often based on hunches, assumptions, or limited personal experiences.

For example, imagine if the faculty developing an ethical reasoning educational program were asked, "Why should this program result in the intended outcome of students acting ethically?" They may state, "We believe lectures on codes of conduct will increase students' knowledge regarding expectations of ethical behavior, and their increased knowledge will increase their ethical behavior." This statement would reflect their program theory (see Figure 1). However, without empirical evidence or established theory to support the link between knowledge of expectations of ethical behavior (intermediate outcome) and ethical behavior (distal outcome), the program theory would be weak.

Figure 1
Logic model depicting the difference between a program with no program theory and a program with weak program theory



In practice, we often observe weak program theory, which limits the use of assessment results to improve ineffective programs (Pope et al., 2019). In fact, we have witnessed rapid development of courses and programs based on hunches or beliefs even though established theory and empirical evidence could have informed course or program development. The most dire situation occurs when existing theory or research provides evidence *against* the hunches or beliefs used to guide program development.

To better integrate articulation of program theory and collection of implementation fidelity data into the assessment of educational programs, we guide readers through a five-step process (see Table 1). The process involves the following steps:

1. Articulate a feasible and malleable distal outcome;
2. Articulate theory- or research-based intermediate (proximal) outcomes;
3. Create intentional, theory- or research-based programming;
4. Collect implementation fidelity data to identify if the research-based programming was implemented;
5. Collect outcomes data to evaluate the effectiveness of the implemented programming.

Answering a series of questions associated with steps 1 to 3 facilitates building the program's logic (see Table 1). The resulting logic model clearly conveys "why" or "how" the programming should impact the distal outcome (see Figure 2). Given strong program theory

Table 1
Five-step process for building and evaluating a theory-based ethical reasoning program

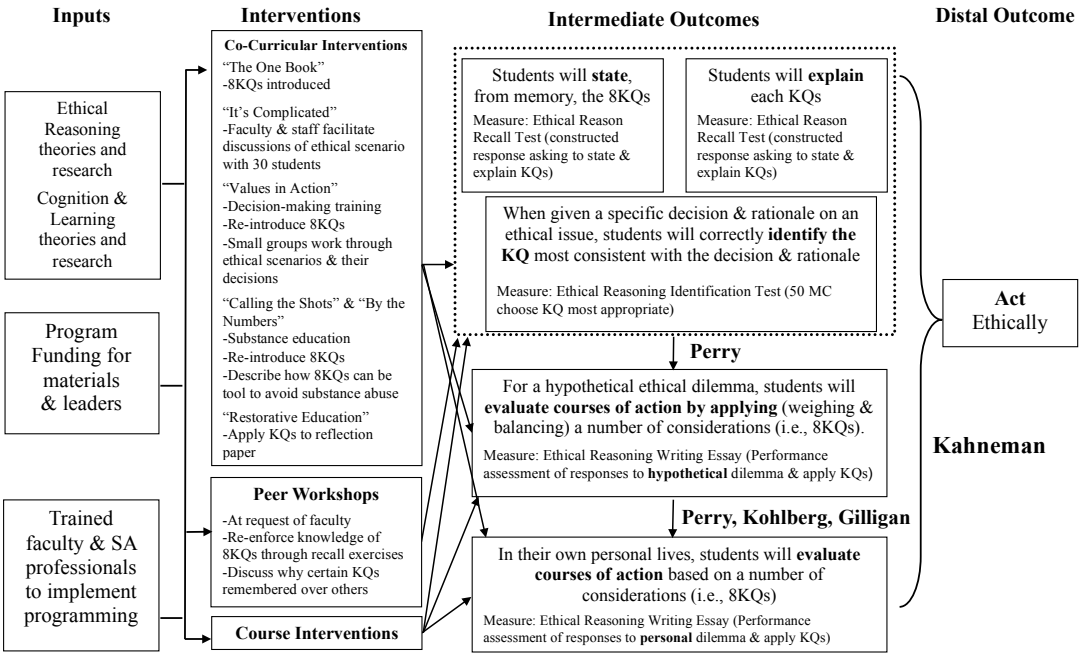
Steps to Articulate & Evaluate Program Theory	Most Important Question(s) to Ask	Ethical Reasoning Example
1. Articulate the Distal Outcome	What problem or outcome needs attention? Is the outcome malleable?	The distal outcome is to significantly increase the frequency of ethical behaviors among students (e.g., students "act ethically").
2. Articulate Theory- or Research-Based Intermediate (Proximal) Student Learning Outcomes (SLOs)	What is the etiology (i.e., what are the causes) of the distal outcome based on current theory and research?	• A deliberative ethical reasoning process is needed to behave ethically. By engaging in a deliberative thought process, students avoid a quick, default, confirmatory decision regarding how to behave (Kahneman, 2011). • To develop this deliberative ethical reasoning process, students must be exposed to multiple considerations associated with an ethical decision or behavior. Traditional students are dualistic thinkers who consider decisions or behaviors as right or wrong (Perry, 1970); thus, we must expose them to multiple considerations. • Upon knowing multiple considerations (e.g., empathy, responsibility), students may tend to favor one consideration over others in most situations (e.g., Gilligan, 1982). Student must be challenged to wrestle with all considerations, which invokes cognitive dissonance and spurs growth in ethical reasoning skills (Gilligan, 1982: Kohlberg, 1981; Perry, 1970). • Given this research, the following intermediate SLOs were specified to influence the distal outcome of behaving ethically: o Students will state, from memory, the 8 Key Questions o Students will explain each Key Question o When given a specific decision and rationale on an ethical issue, students will identify the Key Question most consistent with the decision and rationale o For a hypothetical ethical dilemma, students will evaluate courses of action by applying (weighing & balancing) the Key Questions o In their own personal lives, students will evaluate courses of action based on a number of considerations (i.e., 8KQs)
3. Develop Theory-Based Programming/Curriculum to Impact the Intermediate SLOs	What programming affects the intermediate SLOs based on current theory and research?	• Program consisted of content and activities that foster the encoding, integration, and retention of information about the 8KQs. For example, students experienced at least one knowledge "check point" related to their understanding of the 8KQs, reviewed and refreshed the 8KQs within various case study/dilemma discussions, and mapped or represented the 8KQs in relation to some other work (e.g., disciplinary standards). These types of activities were supported by empirical research from the domain of cognition and learning (Halpern & Hakel, 2003). • Program also consisted of content and activities that utilized case study/dilemma discussions techniques and discipline-specific analysis of ethical cases, decisions, or dilemmas, given research suggested such techniques can promote ethical reasoning development (Bebeau; 1993; Keefer & Ashley, 2001; Wilhelm, 2010). • Program also consisted of numerous opportunities for guided reflection, as research suggested reflection is an important aspect of teaching ethics (Schmidt et al., 2009).
4. Collect Implementation Fidelity Data to Determine if Program Was Delivered as Intended	To what extent did the implemented or delivered program differ from the intended or planned program? Did students actually experience the programming?	• After articulating their program theory, faculty were able to create a fidelity checklist. • Creating the fidelity checklist facilitated the articulation and organization of specific curriculum features. • Reviewing the checklist before each class reminded faculty of the agreed upon and integral program features, guarding against program drift. • The fidelity checklist provided a systematic way to collect fidelity data and thus understand what version of the program students actually experienced.
5. Evaluate Outcomes Data to Inform Inferences about Program Effectiveness and Guide Changes in Program for Learning Improvement	Do assessment results suggest that the programming impacts the intermediate SLOs?	• Outcomes data were collected for the intermediate SLOs to assess change in students' ethical reasoning knowledge (e.g., constructed response, multiple-choice, and performance assessments administered before and after corresponding programming/curriculum). • Outcomes data were not collected for the distal outcome of acting ethically given this behavioral outcome is difficult to collect in real time. However, specification and assessment of theory- and research-based intermediate SLO's led faculty to believe that students achieving these intermediate outcomes are more likely to act ethically. • Fidelity and outcomes data were integrated to make more informed decisions about the program and better understand improvements in students' learning than afforded by outcomes data only.

Note: The program theory incorporates theories and research related to ethical reasoning and moral development to specify and link the proximal intermediate outcomes to the distal outcome of acting ethically (i.e., Step 2). Using theories of learning and cognitive processing, the program theory also explicates how program components should affect the proximal intermediate outcomes (i.e., Step 3).

is articulated, practitioners can then empirically evaluate the theory-based programming (see Table 1, Steps 4 and 5).

Using an ethical reasoning program from our campus, we model a five-step process to articulate strong program theory and assess program effectiveness with regards to learning improvement. Although our example is complex, involving multiple intermediate outcomes and faculty across the institution, the five-step process can be applied to a variety of learning outcomes and educational programming (e.g., Fisher et al., 2014; Gerstner & Finney, 2013; Meixner et al., 2020; Pope et al., 2019).

Figure 2
Example program theory for an ethical reasoning intervention



"Is acting ethically a malleable behavior?" "Can ethical behavior be learned?" If ethical behavior is stable or trait-like, developing programming to try to increase it would be a waste of university resources.

Step 1: Articulate a malleable distal outcome. Creating a theory- or research-based educational program begins by specifying the ultimate or "distal" outcome one hopes to achieve. For the current example, this distal outcome was a result of our institution's Quality Enhancement Plan (QEP) for accreditation through the Southern Association of Colleges and Schools Commission on Colleges (SACSCOC).

Before selecting ethical behavior as the distal outcome, university stakeholders asked: "Is it theoretically possible to impact ethical behavior in a college student population?" "Is acting ethically a *malleable* behavior?" "Can ethical behavior be learned?" If ethical behavior is stable or trait-like, developing programming to try to increase it would be a waste of university resources.

Informed by research, university stakeholders deemed ethical behavior as malleable and they understood intentional instruction would be necessary to build reasoning strategies to influence ethical behavior (Sanchez et al., 2017). For example, Keller (2010) defined ethics as something that can be practiced through "applied methods of rational inquiry to moral problems" (p. 12), suggesting ethical behavior can be impacted by particular approaches. Similarly, research in cognitive psychology provided evidence that many everyday behaviors result from fast, intuitive, or "gut" responses (Kahneman, 2011), which can be interrupted and slowed by the introduction of a prompt or thinking strategy (Ariely, 2013). Reasoning strategies can influence ethical behavior, but these strategies do not develop due to maturation alone. Instead, progression from basic to more advanced stages of reasoning requires effortful development (Kohlberg, 1969; Kohlberg, 1977).

Next, university stakeholders asked, "Given ethical behavior is malleable, can we *impact* it within the time and resource constraints of a traditional, four-year college

experience?" Previous research suggested they could. Since the 1970s, several studies have linked participation in college to ethical reasoning development (King & Mayhew, 2002; Pascarella & Terenzini, 1991; Pascarella & Terenzini, 2005; Rest, 1979; Rest et al., 1986; Rest & Thoma, 1985). Co-curricular and classroom-based experiences have had a significant, but small effect on college students' ethical reasoning, especially for first-year students (Mayhew & Engberg, 2010; Mayhew, et al., 2010).

On our campus, when university stakeholders addressed Step 1, research and theory determined that ethical reasoning was a malleable and feasible outcome to target. Although research suggested that students' ethical reasoning behaviors could be impacted within the context of a college experience (King & Mayhew, 2002; Pascarella & Terenzini, 1991; Pascarella & Terenzini, 2005; Rest, 1979; Rest et al., 1986; Rest & Thoma, 1985), ethical behavior is a complex outcome. University stakeholders anticipated that this outcome may not be realized due to a single program, course, or intervention. Different types of interventions would likely impact different causes of ethical behavior. Thus, during Step 2 of the process, it would be critical for faculty and stakeholders to specify the more proximal, intermediate student learning outcomes that would influence the ultimate, distal outcome of ethical behavior (i.e., "acting ethically").

Research suggested that students need a deliberative ethical reasoning process to behave ethically.

Step 2: Specify theory- or research-based intermediate outcomes. Once the distal outcome of ethical behavior was specified in Step 1, the next step was to consult relevant empirical research and theory to articulate the underlying causes or influences of the behavior. Step 2 is difficult, time consuming, and critically important as it specifies what student characteristics must be influenced to achieve the distal outcome. We provide a detailed description of the process followed on our campus in order to support others engaging in this step.

Faculty tasked with creating the program's curriculum must understand the etiology of acting ethically (West & Aiken, 1997). From this understanding, *intermediate* SLOs were specified (see Table 1 and Figure 2). These intermediate SLOs answered the question: "What do students need to *know*, *feel* or *perceive* (i.e., attitudes), and *do* (i.e., skills) to achieve the distal outcome of acting ethically?" Research suggested that students need a deliberative ethical reasoning process to behave ethically. By engaging in a deliberate ethical reasoning thought process, students can avoid a quick, default, confirmatory decision regarding how to behave (Kahneman, 2011).

Helping students engage in a deliberative thought process can be challenging given students' thought processes are naturally automated, rapid, and rooted in intuitive or "gut" reactions. Thus, students need a strategy or process to help slow their default thinking and instead engage in a deliberative reasoning process (Ariely, 2013). To provide students with such a strategy, stakeholders created the "8 Key Questions" (i.e., the 8KQs) ethical reasoning framework (Sanchez et al., 2017). This deliberative ethical reasoning process prompts students to consider, weigh, and balance the following Key Questions when grappling with an ethical issue or dilemma:

1. Fairness: *How can I (we) act justly, equitably, and balance legitimate interests?*
2. Outcomes: *What possible actions achieve the best short- and long-term outcomes for me and all others?*
3. Responsibilities: *What duties and/or obligations apply?*
4. Character: *What actions help me (us) become my (our) ideal self (selves)?*
5. Liberty: *How do I (we) show respect for personal freedom, autonomy, and consent?*
6. Empathy: *How would I (we) act if I (we) cared about all involved?*
7. Authority: *What do legitimate authorities (e.g., experts, law, my religion/god) expect?*
8. Rights: *What rights, if any, (e.g., innate, legal, social) apply?*

The 8KQs incorporated ideas from the following philosophical perspectives: John Stuart Mill's Utilitarian theory, Kant's natural duties and obligations, Rawls' justice as fairness, Kohlberg's role of authority, Gilligan's role of empathy, and Aristotle's virtuous self (Lehnen & Pyle, 2019).

Committing the 8KQs to memory and being able to explain them were considered necessary (but not sufficient) to acting ethically. That is, the deliberative ethical reasoning process is unpacked in Figure 2 as five intermediate student learning outcomes, with memorizing and explaining the 8KQs being foundational knowledge necessary to engage in the process of ethical reasoning.

Beyond being able to *state* and *explain* the 8KQs, students need to be able to recognize which considerations are being applied by others. As shown in Figure 2, another intermediate outcome involves students being able to *identify* which Key Question is most consistent with a given ethical decision and rationale. Traditional students are dualistic thinkers who tend to consider decisions or behaviors as right or wrong (Perry, 1970). The 8KQs framework intentionally exposes students to multiple considerations associated with an ethical dilemma. Students should understand that a particular ethical dilemma can be associated with any consideration (e.g., Fairness, Authority), and this consideration likely influences one's subsequent behavior.

Upon knowing multiple considerations (i.e., 8KQs), students may tend to favor one consideration over others (Gilligan, 1982). Moreover, stakeholders acknowledged that simply being able to recall, explain, and identify considerations (i.e., KQs) associated with a dilemma (i.e., the first three intermediate SLOs in logic model in Figure 2) would not be sufficient for students to achieve the distal outcome of acting ethically. Thus, two additional, application-focused intermediate SLOs were articulated: students should *evaluate* courses of action by *applying* the 8KQs to hypothetical ethical dilemmas and to personal ethical dilemmas in their own lives (Lehnen & Pyle, 2019). Application of the 8KQ involves weighing the applicability of the considerations raised by each KQ, given the context of the ethical dilemma, and appropriately balancing those considerations to make a conclusion or grapple with a decision. To apply the 8KQs, students must wrestle with all considerations (e.g., 8KQs), which should invoke cognitive dissonance or disequilibrium and spur growth in ethical reasoning skills (Gilligan, 1982; Kohlberg, 1981; Perry, 1970; Schmidt et al., 2009).

The etiology of acting ethically was articulated in Step 2. The specific etiology articulated in Step 2 will vary depending on the distal outcome of interest. Regardless of the outcome, Step 2 involves using theory and research to convey the underlying causes of the desired distal outcome and articulating these in terms of intermediate student learning outcomes.

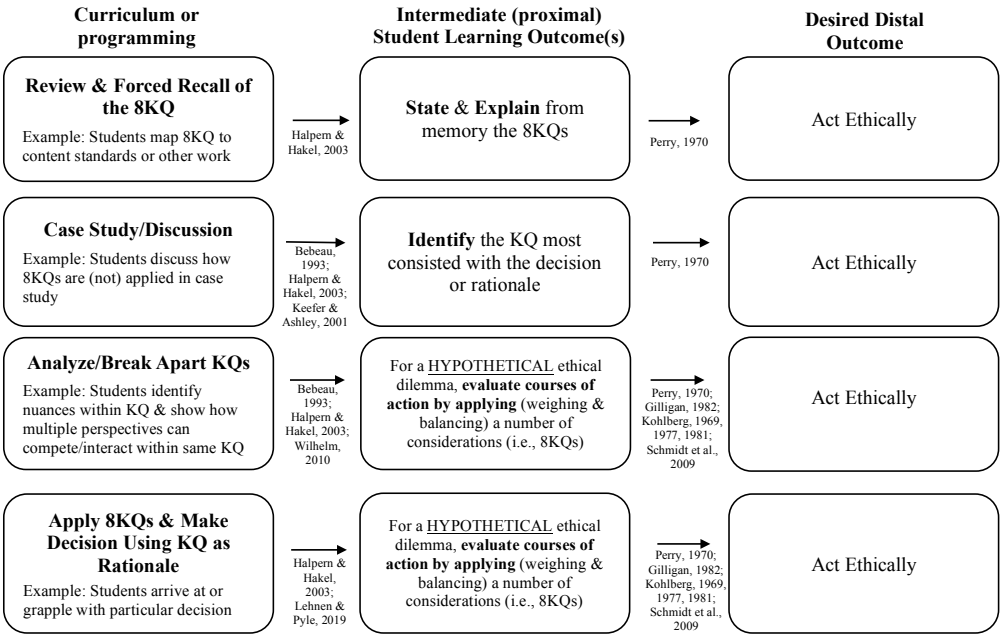
Regardless of the outcome, Step 2 involves using theory and research to convey the underlying causes of the desired distal outcome and articulating these in terms of intermediate student learning outcomes.

For example, using theory and research, stakeholders (including experts in ethical reasoning) indicated that being able to state, explain, and identify the 8KQs facilitates students being able to apply the 8KQs within the contexts of hypothetical and/or personal ethical dilemmas. This application forces students to practice complex thought processes (Lehnen & Pyle, 2019). Being challenged to consider alternative perspectives (i.e., 8KQs) and appropriately weigh and balance these different perspectives (i.e., apply the 8KQs) enables students to act ethically (the desired distal outcome) when they are confronted with ethical dilemmas (Sanchez et al., 2017). These various intermediate SLOs would likely be impacted by different kinds of programming. Thus, during Step 3 of the process, it would be crucial for faculty and stakeholders to specify research-informed programming (i.e., curricula and pedagogies) that would positively influence the intermediate SLOs.

Step 3: Develop theory- or research-based programming aligned to intermediate outcomes. Once the distal outcome of ethical behavior and intermediate SLOs that influence ethical behavior were specified (i.e., Step 1 and 2, respectively), the next step was to determine how to achieve the five intermediate SLOs via programming (e.g., activities, curriculum, pedagogies). At this step, stakeholders asked, "Given achievement of these intermediate outcomes should increase the likelihood of our students acting ethically (i.e., the distal outcome), how can faculty and practitioners *intervene* to support students achieving these intermediate outcomes?" "What *curricular* or *pedagogical strategies* do research or theory suggest may be effective to influence the intermediate outcomes?"

Figure 3

Logic model depicting example curriculum features that were evidence-based resulting in strong program theory



Note: "Hypothetical" ethical dilemmas can be interchanged with "Personal" ethical dilemmas in this logic model.

Just as research and theory informed the articulation of the five intermediate SLOs and the link between them and the distal outcome of ethical behavior, research and theory informed the programming (i.e., curricular, pedagogical components) and linked programming to the intermediate SLOs (see Figure 3). Recall, program theory involves undergirding each arrow in a logic model with research or theory that supports the links (Baldwin et al., 2004). Figure 2 includes several of the interventions built to influence the intermediate SLOs. Some programming was experienced by all students as part of mandatory summer orientation for first-year, entering students. Other programming was experienced by a smaller number of students (e.g., substance abuse education).

Articulating a strong program theory and creating research-informed programming is not sufficient to achieve the SLOs. Students must actually experience the programming in order to achieve the intermediate and distal outcomes.

For the purpose of this paper, we describe the longest intervention— a semester-long ethical reasoning curriculum created by faculty and infused within six cross-disciplinary courses (see Table 2). For the ethical reasoning curriculum, the faculty first examined research that evaluated the effectiveness of particular activities to influence the intermediate SLOs (see Table 1). They consulted research examining best strategies for learning, retention, and future application (e.g., Fink, 2013). They also shared and discussed previous approaches to teaching relevant concepts and identified if those approaches were evidence-based (Smith et al., 2017).

Using this process, faculty co-created a research-informed curriculum that could be implemented across the variety of content domains (see Figure 3). For example, research suggests that practice at retrieval spaced over time promotes long-term retention (Halpern & Hakel, 2003); thus, the faculty designed the ethical reasoning curriculum to include "Review and Forced Recall" activities that would support the encoding and retention of the 8KQs into long-term memory. Students experienced knowledge "check points" of their ability to state and explain the 8KQs. Moreover, encoding of information can be improved when students are asked to "re-represent" information in an alternative format (Halpern & Hakel, 2003). Thus, faculty included a program feature that asked students to map the 8KQs to disciplinary content or other areas of interest (e.g., industry standards, policies of practice, news stories, media).

Research suggests that "Case Study/Discussion" techniques promote ethical reasoning development (Bebeau, 1993). Case-based approaches to teaching ethics have pedagogical utility because they provide opportunities for students to discuss and disagree (Keefer & Ashley, 2001). Thus, the faculty asked students to identify and discuss the

(in)applicability of each of the 8KQs within a given case study or ethical dilemma (see Figure 3). They also asked students to identify aspects of case studies that were compelling in relation to the 8KQs. Furthermore, opportunities for guided reflection are important aspects of teaching ethics (Schmidt et al., 2009). Thus, faculty had students engage in reflection about ethical case studies (e.g., what aspects of case studies were compelling) in formal and informal ways (e.g., oral, written, group, individual).

Table 2
Description of six cross-disciplinary, semester-long courses in which faculty infused theory-based ethical reasoning programming

# of Students	Domain/Discipline	Brief Description of Course	Course Type
77	Health Sciences	Upper level students; Required course for major; Ethics in class title	Lecture
18	Philosophy	Lower level students; General Education Class; Fulfills Cluster 1 requirement; Ethics in class title	Lecture
7	Justice Studies	Upper level students; Elective Course	Seminar; Community Service Learning
7	Integrated Science & Technology	Upper level students; Elective Course	Seminar; Community Service Learning
42	Education	Upper level students; Course for minor	Lecture; Community Service Learning
40	Health Sciences	Upper level students; Required course for major	Lecture

Coupling the implementation fidelity and outcomes assessment data allowed the faculty to understand variability in students' ethical reasoning skills given differences in the extent to which the ethical reasoning programming was implemented with high fidelity.

Research indicates the importance of discipline-specific analysis and examination of ethical issues that students may actually face (Bebeau, 1993; Wilhelm, 2010). Therefore, faculty asked students to "Analyze/Break Apart KQs" relevant to ethical dilemmas within the contexts of their own disciplines (see Figure 3). Students were also asked to grapple with multiple perspectives – within the same KQ – that may compete, interact, or disagree.

Lehnen and Pyle (2019) suggested that students must be challenged to move forward in their ethical reasoning processes and behaviors through interacting with fictional and real-life ethical dilemmas. As students make decisions regarding ethical dilemmas, they will likely experience cognitive dissonance which can spur growth in ethical reasoning skills (Gilligan, 1982; Kohlberg, 1981; Perry, 1970; Schmidt et al., 2009). Therefore, faculty asked students to "Apply the 8KQs and Make Decisions Using KQ as their Rationale" (see Figure 3). For example, students were asked to grapple with a particular ethical decision that someone else made and/or arrive at their own ethical decision using the 8KQ. Faculty also asked students to consider multiple stakeholders and/or perspectives when applying the considerations raised by the 8KQ.

Clearly, the faculty invested substantial time and effort in Step 3 to determine what programming should influence the intermediate SLOs. As outlined in Step 3, development of programming was informed by research, theory, and previous teaching experiences. However, as any instructor knows, articulating a strong program theory and creating research-informed programming is not sufficient to achieve the SLOs. Students must actually *experience* the programming in order to achieve the intermediate and distal outcomes. For students to maximally benefit from research-informed programming, classroom implementation has to be considered (Little & Hahs-Vaugh, 2007). In short, high quality implementation is a necessary aspect of effective programming (Durlak, 2016). Thus, the faculty wanted to empirically evaluate the extent to which the new programming was actually implemented.

Expanding Higher Education Assessment Practice to Include Implementation Fidelity

Implementation fidelity data allow faculty to determine the extent to which the programming *as designed* differs from the programming *as delivered* (Gerstner & Finney, 2013; O'Donnell, 2008; Smith et al., 2017; Smith et al., 2019). Fidelity data allow stakeholders

to better understand the (in)effectiveness of specific features of the educational intervention (Cook & Shadish, 1986) and, in turn, make appropriate modifications (Finney & Smith, 2016). However, traditional outcomes assessment approaches (e.g., Walvoord, 2010) exclude collection of data reflecting the alignment between the *designed* and *delivered* programming. Thus, assessment practice should be expanded to include the collection and use of implementation fidelity data after program theory has been articulated.

Step 4: Collect implementation fidelity data to determine if program was delivered as intended. After faculty co-created the research-based programming aligned with the intermediate outcomes (i.e., Step 3), they needed to determine if the programming was delivered as intended. Thus, they created an implementation fidelity checklist and used it to capture data concerning four aspects of implementation fidelity: 1) whether each program feature was delivered; 2) the quality with which each feature was delivered; 3) the perceived student responsiveness or engagement during a given feature; and 4) the duration of time for each feature (Gerstner & Finney, 2013). See Smith et al. 2017 for more information about the fidelity checklist used by the faculty on our campus.

Implementation fidelity data provided faculty an opportunity to explore the relative effectiveness of specific features of the ethical reasoning programming that they invested a great deal of time and effort co-creating.

Either one or two trained implementation fidelity researchers attended live class sessions throughout the semester and used the checklist to collect fidelity data. Faculty members were asked to review data collected by the fidelity researchers (e.g., to note anything that may have been missed or misrepresented). In addition, for at least three class sessions, each faculty member filled out the checklist as a "self-audit" indication of fidelity (Smith et al. 2017).

In general, the ethical reasoning curriculum was implemented with high fidelity (i.e., strong alignment between planned and experienced programming) because the faculty understood that the influence of their research-informed programming on the SLOs was moderated by implementation fidelity. Articulating a strong program theory enabled the faculty to create a useful implementation fidelity checklist. Creating the fidelity checklist also helped faculty articulate the specific curriculum features. Reviewing the checklist before each class reminded faculty of the agreed upon program features, guarding against program drift. At the end of the semester, faculty commented that using the fidelity checklist added structure to their teaching. The fidelity checklist allowed them to plan their ethical reasoning course materials with greater precision (Smith et al., 2017).

Given the time and resources spent developing the research-informed ethical reasoning programming (i.e., Step 3), the faculty were genuinely excited to assess the extent to which that programming was implemented (i.e., Step 4) and determine if programming was associated with student achievement of the intermediate SLOs (i.e., Step 5). Coupling the implementation fidelity and outcomes assessment data allowed the faculty to understand variability in students' ethical reasoning skills given differences in the extent to which the ethical reasoning programming was implemented with high fidelity. That is, implementation fidelity data provided faculty an opportunity to explore the relative effectiveness of specific features of the ethical reasoning programming that they invested a great deal of time and effort co-creating.

A Call for "Expanded" Assessment Practice in Higher Education

Our experience expanding assessment practice to include strong program theory and implementation fidelity yielded positive results. The combination of program theory, implementation fidelity, and well-aligned outcomes assessment instruments provided:

1. an understanding of *why* students' skills improved over time (e.g., which aspects of the research-informed programming appeared to positively influence students' learning);
2. information to make informed modifications to the programming; and
3. evidence of effective program features that could be shared with colleagues interested in improving similar intermediate or distal learning outcomes.

Had faculty followed a more traditional assessment cycle (e.g., Walvoord, 2010), they *may not* have articulated a research-based program, and they *would not* have collected implementation fidelity data. Had the program theory not been articulated and only the distal outcome of ethical behavior been assessed, how could these limited data be used for program improvement? Had implementation fidelity data not been collected, how could faculty link aspects of programming to improvements in students' learning?

We believe outcomes assessment data have limited utility and thus should not be collected until stakeholders can answer two basic questions: "Why should this programming result in the intended outcome?" (i.e., program theory) and "Was the research-informed programming actually experienced by students?" (i.e., implementation fidelity).

Without strong program theory and implementation fidelity, it is difficult, if not impossible, to determine which intermediate outcomes are achieved and which program features are effective. By specifying how the different program features should (based on research) result in achievement of the intermediate SLOs and how the intermediate SLOs should (based on research) help students progress toward the distal SLO, faculty are able to collect the data necessary to make valid inferences about program effectiveness. Moreover, they can share those results and the new programming with colleagues.

Despite the positive effects of expanding the traditional assessment cycle, there are challenges. For example, a program theory may have been developed to explicate the logic of activities, but the theory was never intentionally communicated (Leeuw, 2003). Thus, new faculty or facilitators may engage in an unnecessary program overhaul because they are not privy to the program's logic. Time must be allocated to create a record of the development of program theory in order to reap the benefits of this difficult, yet critical work. A greater challenge is that faculty or program facilitators may struggle to articulate connections between outcomes and actions (Savaya & Waysman, 2005). Thus, educational interventions are developed (and assessed) without a clear theory or evidence base as a foundation (Bickman, 1987). This challenge stems from the paucity of methods that describe how to do so and lack of training in this domain (Leeuw, 2003; Pope et al., 2019). Similarly, there is a lack of didactic guidance regarding implementation fidelity processes (O'Donnell, 2008; Smith et al., 2017). Collecting implementation fidelity data can be logistically challenging and resource intensive. Moreover, faculty or program facilitators must be willing to have their programming observed, recorded, or otherwise "audited."

Acknowledging these challenges, we urge faculty and practitioners to expand their assessment processes to include the explicit articulation of strong program theory and collection of implementation fidelity data. We have didactically outlined five steps to build and evaluate an evidence-based program that should be effective (see Table 1). We believe outcomes assessment data have limited utility and thus should not be collected until stakeholders can answer two basic questions: "Why should this programming result in the intended outcome?" (i.e., program theory) and "Was the research-informed programming actually experienced by students?" (i.e., implementation fidelity). Although some may find that assertion extreme, it is only after program theory has been articulated that faculty can collect relevant outcomes data. Moreover, valid inferences from outcomes data are contingent on understanding what programming the students actually experienced. This "expanded" assessment practice has great potential to provide better-designed, more effective, research-informed programming. As students have opportunities to experience well-implemented, research-informed programming, their learning should demonstrably improve.

References

Ariely, D. (2013). *The honest truth about dishonesty: How we lie to everyone—especially ourselves*. New York, NY: HarperCollins.

Baldwin, C. K., Hutchinson, S. L., & Magnuson, D. R. (2004). Program theory: A framework for theory-driven programming and evaluation. *Therapeutic Recreation Journal, 38*, 16-31.

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

Banta, T. W., Jones, E. A., & Black, K. E. (2009). *Designing effective assessment: Principles and profiles of good practice*. San Francisco, CA: Jossey-Bass.

Bebeau, M. J. (1993). Designing an outcome-based ethics curriculum for professional education: Strategies and evidence. *Journal of Moral Education, 22*, 313-326.

Bickman, L. (1987). The functions of program theory. *New Directions for Program Evaluation, 33*, 5-18.

Coates, H. (2016). Assessing student learning outcomes internationally: Insights and frontiers. *Assessment & Evaluation in Higher Education, 41*, 662–676.

Cook, T. D., & Shadish, W. R. (1986). Program evaluation: The worldly science. *Annual Review of Psychology, 37*, 193-232.

Durlak, J. A. (2016) Programme implementation in social and emotional learning: Basic issues and research findings. *Cambridge Journal of Education, 46*, 333-345.

Fink. L. D. (2003). *Creating significant learning experiences: An integrated approach to designing college courses*. San Francisco, CA: Jossey-Bass.

Finney, S. J., & Horst, S. J. (2019a). Standards, standards, standards: Mapping professional standards for outcomes assessment to assessment practice. *Journal of Student Affairs Research and Practice, 56*, 310-325.

Finney, S. J., & Horst, S. J. (2019b). The status of assessment, evaluation, and research in student affairs. In V. L. Wise & Z. Davenport (Eds.), *Student affairs assessment, evaluation, and research: A guidebook for graduate students and new professionals* (pp. 3-19). Springfield, IL: Charles Thomas Publisher.

Finney, S. J., & Smith, K. L. (2016). *Ignorance is not bliss: Implementation fidelity and learning improvement*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA). https://www.learningoutcomesassessment.org/wp-content/uploads/2019/08/Viewpoint-FinneySmith.pdf

Fisher, R., Smith, K. L., Finney, S. J., & Pinder, K. (2014). The importance of implementation fidelity data for evaluating program effectiveness. *About Campus, 19*, 28-32.

Fulcher, K. H., & Prendergast, C. O. (2019). Lots of assessment, little improvement? How to fix the broken system. In S. P. Hundley & S. Kahn (Eds.), *Trends in assessment: Ideas, opportunities, and issues for higher education*. Sterling, VA: Stylus.

Fulcher K. H., Smith, K. L., Sanchez, E. R. H., Ames, A. J., & Meixner, C. (2017). Return of the pig: Standards for learning improvement. *Research & Practice in Assessment, 11*, 10-27.

Gerstner, J. J., & Finney, S. J. (2013). Measuring the implementation fidelity of student affairs programs: A critical component of the outcomes assessment cycle. *Research & Practice in Assessment, 8*, 15-28.

Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.

Halpern, D. F., & Hakel, M. D. (2003). Applying the science of learning. *Change: The Magazine of Higher Learning, 35*, 36-41.

Jankowski, N. A., Timmer, J. D., Kinzie, J., & Kuh, G. D. (2018). *Assessment that matters: Trending toward practices that document authentic student learning*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus, and Giroux.

Keefer, M., & Ashley, K. (2001). Case-based approaches to professional ethics: A systematic comparison of students' and ethicists' moral reasoning. *Journal of Moral Education, 30*(4), 377-398.

Keller, D. R. (2010). An introduction to ethics for teaching. *Teaching Ethics, 11*, 1-54.

King, P. M., & Mayhew, M. J. (2002). Moral judgment development in higher education: Insights from the defining issues test. *Journal of Moral Education, 31*, 247-270.

Kohlberg, L. (1969). Stage and sequence: The cognitive approach to socialization. In D.A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347-480). Chicago, IL: Rand McNally.

Kohlberg, L. (1977). Implications of moral stages for adult education. *Religious Education, 72*, 183-201.

Kohlberg, L. (1981). *Essays on moral development: The philosophy of moral development* (Vol. 1). San Francisco, CA: Harper & Row.

Kushimoto, T. (2010). Outcomes assessment and its role in self-reviews of undergraduate education: In the context of Japanese higher education reforms since the 1990s. *Higher Education, 59*, 589-598.

Leeuw, F. L. (2003). Reconstructing program theories: Methods available and problems to be solved. *American Journal of Evaluation, 24*(1), 5-20.

Lehnen, A. R., & Pyle, L. K. (2019). Ethical and moral reasoning development across campus. In K. Branch, J. S. Hart-Steffes, & C. M. Wilson (Eds.), *Applying student development theories holistically: Exemplar programming in higher education* (pp. 79-93). New York, NY: Routledge.

Leathwood, C., & Phillips, D. (2000). Developing curriculum evaluation research in higher education: Process, politics and practicalities. *Higher Education, 40*, 313-330.

Little, M. E, & Hahs-Vaughn, D. L. (2007). The implementation of content enhancement routines for improved content literacy for middle and secondary social studies students. *Journal of Personnel Evaluation in Education, 20*, 261-280.

Mathers, C.E., Finney, S. J., & Hathcoat, J. (2018). Student learning in higher education: A longitudinal analysis and faculty discussion. *Assessment and Evaluation in Higher Education, 43*, 1211-1227.

Mayhew, M. J., & Engberg, M. E. (2010). Diversity and moral reasoning: How negative diverse peer interactions affect the development of moral reasoning in undergraduate students. *The Journal of Higher Education, 81*(4), 459-488.

Mayhew, M. J., Seifert, T. A., & Pascarella, E. T. (2010). A multi-institutional assessment of moral reasoning development among first-year students. *The Review of Higher Education, 33*(3), 357–390.

Meixner, C., Pope, A. M., & Horst, S. J. (2020). Implementation fidelity in the classroom: Exploring the alignment of pedagogy to learning outcomes. *The Journal of Faculty Development, 34*, 15-22.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78*, 33-84.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students*. San Francisco, CA: Jossey-Bass.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students*. San Francisco, CA: Jossey-Bass.

Perry, W. G. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York, NY: Holt, Rinehart, & Winston.

Pope, A., Finney, S. J., & Bare, A. (2019). The essential role of program theory: Fostering theory-driven practice and high-quality outcomes assessment in student affairs. *Research & Practice in Assessment, 14*, 5-17.

Rest, J. R. (1979). *Development in judging moral issues*. Minneapolis, MN: University of Minnesota Press.

Rest, J. R., Deemer, D., Barnett, R., Spickelmier, J., & Volker, J. (1986). Life experiences and developmental pathways. In J. Rest (Ed.), *Moral development: Advances in research and theory* (p. 28-58). New York, NY: Praeger.

Rest, J. R., & Thoma, S. (1985). Relationship of moral judgment development to formal education, *Developmental Psychology, 21*, 709-714.

Sanchez, E. R. H., Fulcher, K. H., Smith, K. L., Ames, A. J., & Hawk, W. J. (2017, March-April). Defining, teaching, and assessing ethical reasoning in action. *Change: The Magazine of Higher Learning, 49*, 30-36.

Savaya, R., & Waysman, M. (2005). The logic model: A tool for incorporating theory in development and evaluation of programs. *Administration in Social Work, 29*(2), 85-103.

Schmidt, C., McAdams, C., & Foster, V. (2009). Promoting the moral reasoning of undergraduate business students through a deliberate psychological education-based classroom intervention. *Journal of Moral Education, 38*, 315-334.

Smith, K. L., Finney, S. J., & Fulcher, K. H. (2017). Actionable steps for engaging assessment practitioners and faculty in implementation fidelity research. *Research & Practice in Assessment, 12*, 71-86.

Smith, K. L., Finney, S. J., & Fulcher, K. H. (2019). Connecting assessment practices with curricula and pedagogy via implementation fidelity data. *Assessment and Evaluation in Higher Education, 44*, 263–282.

Smith, K. L., Good, M. R., & Jankowski, N. (2018). Considerations and resources for the learning improvement facilitator. *Research & Practice in Assessment, 13*, 20-26.

U.S. Department of Education. (2006). *A test of leadership: Charting the future of U.S. higher education* (ED Pubs No. ED002591P). Retrieved from: http://www2.ed.gov/about/bdscomm/list/hiedfuture/reports.html

Walvoord, B. E. (2010). *Assessment clear and simple: A practical guide for institutions, departments, and general education* (2nd ed.). San Francisco, CA: Jossey-Bass Publishers.

West, S., & Aiken, L. (1997). Toward understanding individual effects in multicomponent prevention programs: Design and analysis strategies. In K. Bryant, M. Windle, & S. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 167–209). Washington, DC: American Psychological Association.

Wilhelm, W. (2010). Ethical reasoning instruction in non-ethics business courses: A nonintrusive approach. *The Delta Pi Epsilon Journal, 52*, 152-167.

Abstract

Traditional classroom assessment practice often leaves students out of the conversation, exacerbating the unequal power distribution in the classroom. Viewing classrooms as autonomy-inhibiting is known to influence students' psychosocial wellbeing as well as their academic achievement. This is especially relevant in STEM fields where marginalized populations experience disproportionate rates of attrition and success. The current paper describes the use of a simple, piloted classroom intervention with the intention of incorporating student voice in the co-creation of assessment criteria, relative to participation evaluation. Undergraduate students in a STEM research methods design course took part in the intervention. Pre- and post-tests were administered to understand what effects, if any, the intervention had on student perceptions of power and attitudes towards assessment. Perceptions of power significantly increased from pre- to post-test, and qualitative feedback from the intervention were overwhelmingly positive. Limitations and suggestions for future research are also discussed.

AUTHORS

Manisha Kaur Chase, M.A.
University of California Los Angeles

Student Voice in STEM Classroom Assessment Practice: A Pilot Intervention

Traditional classroom assessment practice is rarely known to involve students in the conception of assessment purpose or design (Falchikov, 2004). It follows that such an exclusion of student voice in assessment exacerbates the unequal power dynamics students experience in the classroom (McCroskey & Richmond, 1983; Sidky, 2017). This is especially important in Science, Technology, Engineering and Math (STEM) fields where marginalized populations experience disproportionate rates of attrition and greater barriers to success. To address these dynamics and to afford students more agency in their learning, recent higher education trends, particularly in STEM fields, show a shift towards student-centered pedagogy including the flipped classroom model and inquiry-based instruction, among others. Researchers argue, however, that assessment has been "neglected" in the pursuit of these strategies (Wanner & Palmer, 2015). The current paper describes the use of a simple, piloted classroom intervention with the intention of incorporating student voice in the co-creation of assessment criteria, relative to participation evaluation, and the larger goal of addressing some of the power imbalances and inequities present.

CORRESPONDENCE

Email
mkc012@g.ucla.edu

Student Voice & Co-Creation

"Student voice" is referred to here as "efforts that strive to redefine the role of students in educational research and reform" (Cook-Sather, 2006). The term was originally defined in the context of K-12 education but has also been applied to higher education (Brooman et al., 2015; Monsen & Cook, 2017). While attempts to involve students in

educational decisions at the university level is not a recent phenomenon, there has been as of late, a push to differentiate the types and respective consequences of current methods. For example, Bovill and Woolmer (2019) made a distinction between instructor and student "co-creation *of* the curriculum" versus "co-creation *in* the curriculum." Co-creation of the curriculum refers to that which occurs prior to the implementation of practice and often only includes fewer student voices in the process. In contrast, co-creation in the curriculum occurs during implementation of practice and typically engages a whole classroom of student voices. The latter strategy is least-often used in research and practice in higher education for various logistical challenges such as classroom size and time constraints. This is evident in the assessment realm, too, as studies that attempted to bring students into the assessment conversation revealed decisions about assessment practice having been made in isolation or prior to the beginning of a course (Pretorius et al., 2017; Wanner & Palmer, 2015).

The aim of incorporating student voice into classroom matters—specifically that of assessment—is to trouble the notion of power norms and bring marginalized voices to the fore.

The aim of incorporating student voice into classroom matters—specifically that of assessment—is to trouble the notion of power norms and bring marginalized voices to the fore; as demonstrated by previous work that specifically aimed to involve students in assessment practice (Bovill, 2020; Guberman, 2020; Deeley & Brown, 2014). However, in the case of co-creation of the curriculum, it is only the select students (typically those who have already taken a course, for example) that are privileged with the opportunity to have their voice heard as a result of their retrospective experience. Thus, while one could argue that student voice is being considered, the voices of those students *currently* in a given course are not consulted, which does little to challenge the power dynamic from their perspective. In this way, there is a need for co-creation in the curriculum that actively seeks to acknowledge student voice as their experiences in the classroom take place in order to provide enhanced inclusion and a more equitable learning environment. In doing so, such an approach may also serve to mitigate the power disparity students perceive in the classroom.

STEM Context

The current study is not only concerned with the intersection of student voice and power in classroom assessment practice, but more specifically, in the STEM classroom. STEM disciplines are known to be "cutthroat" in nature (McGee, 2016) with "individualistic weed-out culture" (Daily et al., 2007) that are largely grounded in white, middle-class, masculine norms (Fabert et al., 2011). Retention in STEM is low across the board but known to disparately affect ethnic and gender minoritized groups (Chen & Soldner, 2013).

At the heart of many explanatory factors that attempt to narrow down the origins of this phenomenon (including a lack of belongingness, loneliness, and powerlessness) are assessment practices and grades which not only neglect to include these marginalized voices in their conception, but also expose the "gatekeeping" function of these assessments and the larger gatekeeping function of STEM as a whole (Committee, 2016). This perceived distance between students' respective identities and STEM content is reflected in assessment practices that are not only traditional in nature (quizzes and tests relying heavily on memorization techniques), but also "devoid of deep connection" to real life given their focus on isolated facts (Martin-Hansen, 2018; Momsen et al., 2010). In other words, despite the increasing diversity of undergraduate student populations, STEM classrooms often operate on implicit curriculum that reinforce the perceived objectivity of the disciplines themselves. Therefore, STEM is a context that may especially benefit from considering the need for student voice and co-creation in classroom assessment, but it is also one that requires a unique consideration of its culture, including hegemonic norms and the way those norms are perceived and acted upon by both instructors and students. The current paper and pilot intervention answers the call for inviting student voice and co-creation in the STEM classroom towards developing student autonomy (Bovill & Woolmer, 2019; Evans & Boucher, 2015), gauging psychosocial aspects of a classroom from the student perspective (Drewes et al., 2019), and engaging in assessment that explicitly acknowledges power stratification in the classroom and seeks to empower students (Bain, 2010).

Theoretical Framework

Self-determination theory (SDT) is used here to highlight the need for examining autonomy development in the assessment realm. SDT is a theory of motivation which dichotomizes motivation as autonomous versus controlled (Deci & Ryan, 2008; Deci & Ryan, 1985). Autonomous (or intrinsic) motivation occurs when an individual experiences volition, through internal factors free from outside pressure or reward, versus controlled (or extrinsic) motivation in which an individual's actions are a result of external rewards or punishments.

In education, SDT argues that student motivation to learn can be explained by how well teachers encourage individual growth– including via autonomy. When an educational environment is perceived as having no room for control, self-determination—and consequently, motivation—experience decline (Deci et al., 1989). With a focus on choice and control, SDT situates the way a particular context enables or inhibits student voice/co-creation and its subsequent effects on individual autonomy (Ryan & Niemiec, 2009).

Student autonomy relative to assessment would require students having the opportunity to "take charge" (Holec, 1981) of those assessments that contribute to and evaluate their learning. In fact, an SDT approach to assessment would "actively empower and support change from within" (Ryan & Weinstein, 2009). However, the current state of assessment affairs—particularly those summative in nature—are predominantly conceived and implemented by instructors. Without experience in the assessment realm, students are left with a hierarchy that perpetuates student dependency on teacher judgement (e.g., grades) which may stifle student autonomy and exaggerate the power stratification (Sadler, 1989). However, by actively incorporating student voice in the classroom and increasing perception of choice, student autonomy development may flourish. Greater sense of student autonomy has been linked with greater internalized motivation, as well as better academic outcomes (Black & Deci, 2000; Chirkov & Ryan, 2001). In order to reap the benefits of empowering students, it becomes necessary not only to understand but also to address the power dynamics in a classroom—including how the dynamics are perceived, produced, and sustained—in attempts to engage students in the learning process in ways that bolster their motivation and sense of autonomy.

Current Practice

In other words, despite the increasing diversity of undergraduate student populations, STEM classrooms often operate on implicit curriculum that reinforce the perceived objectivity of the disciplines themselves.

Current research demonstrates a variety of ways students have been involved in flexible, student-centered assessment practices. For example, a study by Pacharn et al. (2013) examined the effects of allowing undergraduate students the flexibility to allocate weights to the various assessments in their course. The full-flexible group (allowed to adjust weights until the end of the semester) reported significant increases in motivation, attitudes towards motivation, and higher academic outcomes than their early-flexible (only allowed to choose their weights at the beginning of the semester) and control group peers. In another example of flexible assessment, undergraduate students were given the opportunity to choose between multiple grading schemes for their course (Rideout, 2018). While choice of scheme had no significant correlation with final course outcomes, 79% of students reported being satisfied with having the choice.

These studies do provide an element of student choice in the assessment realm but have faltered in acknowledging student voice and supporting true student autonomy in the following ways. Firstly, assessment interventions that allow students to alter grading schemes and weightings serve only to reinforce grades and suggest that the conversation about assessment be limited to these quantifiers, rather than about the learning process as a whole. Moreover, in other examples where students were given the opportunity to choose a topic within an assignment or project (Bullen, 2012; Vandiver & Walsh, 2010), or provide feedback on a rubric (Fletcher & Shaw, 2012) they did so within the confines of an existing assessment format. In other words, students were not involved in the discourse on what purposes the assessments would serve, or which assessments might best serve said purposes. Students were involved only *after* these decisions had been determined. Finally, it should be noted that studies in this area rarely probe students' perceptions of power as a result of being involved in assessment dialogue. In one example of a partnership approach

to assessment, researchers found students expressed increased motivation and engagement as a result of being part of a more "democratic" assessment process (Deeley & Bovill, 2017). These data, however, were collected from qualitative student open-ended responses, and were not prefaced with a baseline measure of students' perceptions prior to being engaged in the process.

Participation Evaluation

Engaging student voice in assessment practice may be an intimidating and confusing activity for students (Hewitt-Taylor, 2001; Monsen et al., 2017). Additionally, while dated, there are concerns that students do not have sufficient content knowledge in order to be involved in assessment practice (Falchikov, 2004; French et al., 1959). Thus, the current study suggests involving student voice in an area of evaluation that does not require content mastery: classroom participation.

In a survey of one urban university, 93% of all courses included participation as part of overall course grades (Bean & Peterson, 1998). A survey of 520 instructors at another large, state university revealed 82% of faculty reportedly including participation in their syllabus, with only 25% of these professors actually providing criteria and grades for participation (Rogers, 2013). Why were grades not formally assigned to participation? One professor noted, "'I believe that different students learn in different ways, and forcing quiet students to talk in class is obnoxious and likely to be counterproductive, e.g., superficial participation just for points'" (p. 18). While on the surface this appears to be a considerate reflection from the instructor's point of view, their comment serves to ignore the student voice regarding what good participation may look like (i.e., in the case of the "quiet student"), and rests on an esoteric, normative assumption that good participation is represented solely by speaking up in class. By dismissing the evaluation altogether, the professor missed an opportunity to engage student voice through co-creation in assessment, in order to develop criteria that may be representative of the diversity of students and their respective needs and preferences. Such a critical perspective—that criteria must take into account minoritized students, cultural conflicts, and issues of representation and power in the classroom, relative to participation evaluation—has also been cited in the literature (Meyer & Hunt, 2011; White, 2011).

[This] study set out to explore how an intervention that highlights undergraduate STEM students' voices through co-creation in participation criteria might affect student perceptions of autonomy support (power) and attitudes towards assessment.

Given the prevalence of participation evaluation in course grades and the discord between its mention and its actual evaluation, I argue this is an appropriate area to begin incorporating student voice in assessment practice.

While research has shown previous attempts to include students in the creation of participation criteria, these attempts have failed to have students come up with the operationalization of participation skills and grading, and have not yet shown the impact on student perceptions of psychosocial outcomes (Dancer & Kamvounias, 2005). Thus, the current pilot study set out to explore how an intervention that highlights undergraduate STEM students' voices through co-creation in participation criteria might affect student perceptions of autonomy support (power) and attitudes towards assessment.

Method

This quasi-experimental pilot study employed a pretest-posttest design to uncover the potential effects of an assessment intervention on student perceptions of their classroom. An informed consent waiver was distributed to all students outlining their participation in the study. IRB approval was obtained in order to disguise the intervention's true purpose and risk biasing survey responses wherein students were told the study was vaguely aimed at understanding student perceptions of their classroom. Following the study, students were debriefed as to the true intentions of the study, including explicitly addressing the intervention and its relation to perceptions of power and attitudes towards assessment specifically.

Participants and Context

The pilot intervention was carried out with a total of 21 undergraduate students (ages 18-22) from an upper-division, STEM, research methods design lab course at a large,

public university. The sample consisted of seven self-identified Caucasian/White students, four each self-identifying as Latinx/Hispanic, East Asian, and Multiethnic, respectively, while the remaining students self-identified as Southeast Asian or Middle Eastern. Three-quarters of the sample self-identified as female while the remaining self-identified as male. Three students identified as being transfers, two reported international student status, and seven identified as first-generation college students. Only one student reported having a learning accommodation. Overall, these demographics are reflective of the specific STEM major at this institution. The course itself was comprised of a weekly asynchronous lecture taught by the instructor of record while the lab course met synchronously with graduate Teaching Assistants (TA)—twice weekly for two hours each over the course of an academic quarter (10-week period). Given the ongoing global COVID-19 pandemic at the time of this course, the usual in-person curriculum was adapted for online instruction (including the synchronous lab discussions).

Course grades were a compilation of the lecture exams (40%) and the lab assignments and activities (60%). Participation consisted of 20% of students' lab grades, which included a day-to-day evaluation from the TA in addition to an end of quarter peer-assessment relative to a group project implemented in the course. The current intervention focused on the criteria which informed the day-to-day evaluation portion of students' participation evaluation. Typically, the criteria for what constitutes good participation is not mandated from the instructional team and thus, left up to the discretion of the TA. In this pilot study, the researcher capitalized on this flexibility in an attempt to bring students into the assessment conversation.

Measures

Power was operationalized as students' perceptions of autonomy support from their instructor in addition to their perception of having voice in the classroom. The 6-item "Learning Climate Questionnaire" [(LCQ); Williams & Deci, 1996] was adapted for the purpose of this study [α =.65, p=.001; acceptable reliability as compared to $\alpha \geq .60$ (van Griethuijsen, 2015)]. Participants were prompted to "think about the way you are assessed by your TA and respond to the following prompts in regards to that assessment experience." Sample items included: "I feel that my TA provides me choices and options" and "My TA conveyed confidence in my ability to develop assessment criteria." Responses fell on a Likert scale from 1-7 with 1 representing "Strongly Disagree" and 7 representing "Strongly Agree." Item responses were aggregated into a single perception of power score for each participant, where higher aggregated scores suggested increased perceived autonomy support/power.

A final purpose of the intervention was to provide students a holistic experience—from the very beginning of determination of purpose to the 'end result' of grading itself—of assessment in the classroom.

Meanwhile, student attitudes toward assessment was operationalized as students' preference and beliefs regarding assessment in their classroom. A 5-item version adapted from the "Attitudes towards Grading System" scale (Pacharn et al., 2013) was used to gauge student attitudes (α =.24, p >.05). Sample items included: "I liked how the grading scheme employed in this course, with respect to participation, was determined" and "I believe that allowing students to participate in designing the grading scheme (e.g., in relation to participation) in a course wastes students' time that could be better spent working on the course material." Participants responded on a 7-point Likert scale with 1 indicating "Strongly Disagree" and 7 indicating "Strongly Agree." Once again, item responses were aggregated into a single attitude towards assessment score for each participant.

Finally, in order to understand the qualitative experience of the intervention, a survey administered half-way through the quarter consisted of a free response where students had the opportunity to describe how the experience of being involved in assessment development made them feel, what effect it had on their perceptions of the classroom/instructor, what they enjoyed about the experience, and what might be used to improve the intervention. This provided anecdotal data on students' experience of and suggestions to improve the intervention.

Procedure

The overall aim of the pilot intervention—as outlined in detail below—was to involve student voice in classroom assessment practice. More specifically, the intervention sought

to achieve the following: meaningfully engage student voice in the assessment development process through the co-creation of participation evaluation criteria and provide students an opportunity to stray away from the historical dependence on instructors for assessment evaluation. Additionally, as a result of having to create the criteria in addition to apply it via self-assessment, a final purpose of the intervention was to provide students a *holistic* experience—from the very beginning of determination of purpose to the 'end result' of grading itself—of assessment in the classroom.

The pre-survey was administered on the first day of lab during Week 1, followed by the qualitative survey gauging the intervention process at Week 5, and finally, the post-survey at Week 10.

The Intervention

The proposed intervention took place on the first day of class as the Teaching Assistant went over the syllabus policies of the lab. Following the pre-survey, the TA prefaced the intervention by discussing the challenges of assessment in school contexts. More specifically, the challenge of attempting to measure something, unlike weight or height, that is not tangible. The TA cited a history of researchers, policy makers, professors, etc. working to hone assessment practices to make them fair, valid for their outset purpose, and reliable. However, the TA noted this iterative process has often failed to incorporate student voice in what is classified as important. The TA then expressed that in the current context, they wanted to give that opportunity to students such that they may co-create meaning of one aspect of assessment in the course: participation.

The final five skills included: written communication, verbal communication, engagement, critical thinking, and teamwork.

In guiding students to think about assessment purposes at large, the TA first asked the class why participation may be a part of their grade (when it seemingly has nothing to do with STEM content). Probing questions included: "What might participation (in all its forms) be representative of? What skills might we be assessing when it comes to the various aspects of assessments?" Essentially, the TA asked students to think relative to *this specific course* what skills students valued and wanted to foster in their space. Using the web-based response site, *Mentimeter*, students were asked to record three larger skills they believe participation represents. These were then generated into one cohesive word cloud (see Figure 1). Using the word cloud, the TA engaged students in a discussion of consolidating these into a handful of larger skills participation would be representative of in the course. The final five skills (read here as purposes of participation evaluation in this context) included: written communication, verbal communication, engagement, critical thinking, and teamwork.

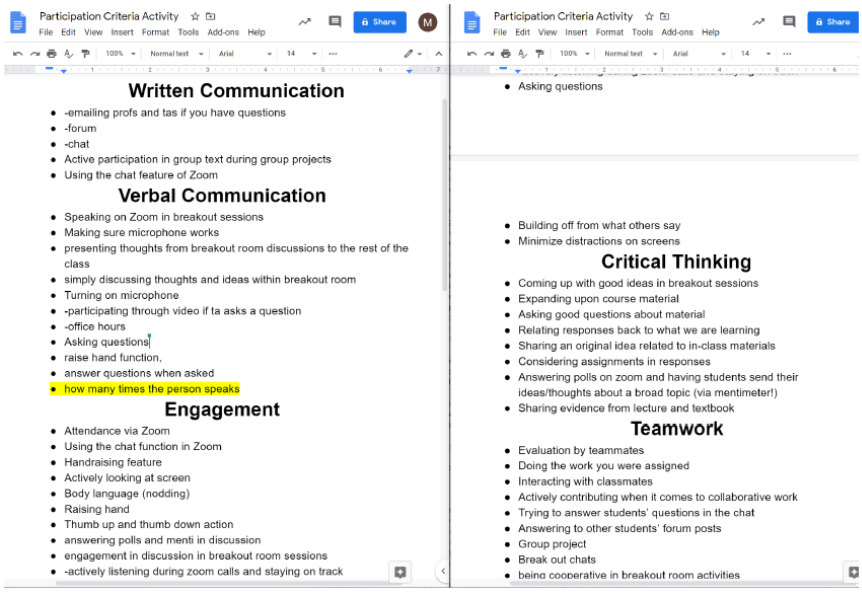
Once these skills were established, the TA split students up into smaller groups to discuss what concrete behaviors might be representative of each of those skills (i.e., a skill of being respectful in the classroom might have a concrete behavior of not being distracted via cell phone or laptop use). Each group elected a scribe who transferred the group notes into a shared Google Doc (see Figure 2). The TA then had students return from their groups and take a few minutes to review the criteria their peers had constructed. This was followed by a facilitation of how students felt about the criteria in general, and if there were any criterion students would like to make more specific, or perhaps, remove.

In this iteration of the intervention a student did in fact raise a concern. Under the skill "verbal communication," one group had suggested, "how many times a student speaks," as a potential criterion to be used. The student contested that perhaps, particularly in an online format, asking students to speak may be uncomfortable for some and that additionally, quantifying verbal communication might lead to students "speaking for the sake of it" while not lending any meaningful contribution to discussion. The TA then asked if there were any counterpoints to the concern raised, and an anonymous poll was employed in which the students unanimously voted to expel that criterion.

Following this process, the TA created a polished version of the criteria and noted that this is what would be used to evaluate student participation in lab (see Figure 3). In order to scaffold the assessment development experience for students, a scale for grading was suggested for the criteria created. The TA explained how the criteria would align with the three

Figure 1
Sample Student Word Cloud of Meta-skills Represented by Participation

Figure 2
Sample Working Google Document of Purposes and Criteria



participation points students could earn for each lab session (a combination of the quantity and quality of student engagement). To avoid confusion, it was cautioned that students were not expected to engage in *all* criteria. Overall, the process took approximately 20 minutes. This concluded the first part of the intervention.

The second part of the intervention took place during Week 5 of the quarter where students were reminded of the criteria and had an opportunity to engage in self-assessment. First, students were administered the qualitative survey which probed affective feelings about the intervention in addition to feedback on the process. Thereafter, students were asked to qualitatively self-assess how they feel they had lived up to the participation criteria they developed. Finally, space was provided for students to indicate any criterion they would like to change or add given their experience in the first half of the course. Following class, the students were notified by the TA of the score they had accumulated in the first half of the quarter according to the TA's evaluation using the criteria. This provided one way in which students could understand how the criteria they developed had resulted in their actual participation grade as well as an opportunity to show that the criteria they created were in fact

Figure 3
Sample Finalized Criteria

Participation Criteria for [redacted]

Written Communication	Verbal Communication	Engagement	Critical Thinking	Teamwork
• Emailing the professor or TAs if you have questions • Using the CCLE forum to ask questions/pose answers • Using the chat function on Zoom • Active participation in group text during group projects	• Using the raise hand function or simply raising hand • Sharing thoughts, ideas, and questions verbally in Zoom breakout sessions/main room • Making sure microphone works • Attending office hours • Providing answers to questions posed by TA and classmates	• Attendance via Zoom • Using the chat function, thumbs up, or raising hand in Zoom • Using body language such as actively looking at screen, nodding, thumbs up, or raising hand • Answering polls and Mentimeter questions • Actively listening • Asking questions • Building off of what the TA or peers say • Minimize distractions on screens	• Coming up with good ideas in breakout sessions/main room • Expanding upon course material, including asking questions • Relating responses back to what we are learning in lecture • Sharing an original idea related to in-class materials • Considering assignments in responses • Sharing evidence from lecture and textbook (and potentially, real life)	• Evaluation by teammates • Doing the work/actively contributing to what you were assigned by your group • Interacting with classmates during breakout sessions • Staying on task within groups • Trying to answer peers' questions in the chat and/or forum posts

being used by the TA. Sixty-two percent of the class at this time reported being satisfied with the criteria, while the remaining students had comments relative to the quantity of criteria. In the words of one student: it was "hard to narrow down the main points." The TA addressed these concerns citing the need for inclusivity such that the breadth of criteria was meant to cater to individual strengths in participating, rather than being a punitive system that required students to participate using all suggested criteria. Comments and suggestions were welcomed but appeared to have been satisfied with the aforementioned explanation. Finally, students were given the post-survey in Week 10.

Results

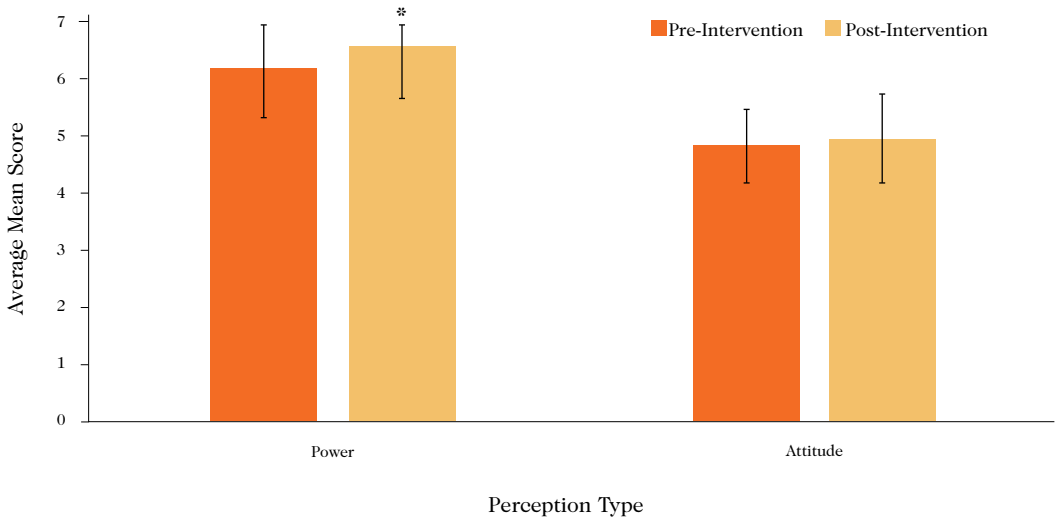
Figure 4 presents the average aggregate score of perception of power and attitudes towards assessment pre- and post-intervention, respectively. The average perception of power during the pre-test (M=6.02, SD=0.79) was lower than the average reported in the post-test (M=6.44, SD =0.88). Similarly, the average attitude towards assessment was lower in the pre-test (M=4.74, SD =0.65) than that of the post-test (M=4.88, SD =0.79). A Shapiro-Wilk test for normality revealed a non-normal distribution of both pre- and post-test power scores (p=.005; p=.000) and a normal distribution for both pre- and post-test attitude scores, respectively (p=.38; p=.19).

The average perception of power during the pre-test was lower than the average reported in the post-test.

In order to compare within-subject differences, a related-samples Wilcoxon Signed Rank Test for nonparametric data was conducted for power scores, and a paired-samples t-test for attitudes towards assessment were conducted to reveal any significant changes in mean perceptions reported from the pre-post surveys. These tests revealed a significant difference in perceptions of power from pre-survey to post-survey (Z=91, p=.015), where student perceptions of their own power in the classroom were higher, or closer to "Agree," on the post-test compared to the pre-test. The effect size for this analysis (d =-1.12) was found to exceed Cohen's (1998) assumption for a large effect (d =0.80). Finally, no significant difference for attitudes toward assessment were found from pre- to post-survey (t(20) =-.69, p=.50; d =-0.15).

The open-ended responses were analyzed using an inductive, open coding process in order to allow similar themes to arrive from data itself, rather than employing a pre-conceived, deductive coding scheme to student experience. The first round focused on those experiences that students appeared to share during the process, the second round consisted of consolidating those themes, and the final round searched for exceptions to the themes.

Figure 4
Perception of Power and Attitudes Toward Assessment Scores Pre and Post Intervention



When asked about how students felt they had lived up to the criteria, many responded they felt they had adequately met a fair number of criteria—particularly those criteria that played to their strengths. One student felt they had not met the criteria, citing difficulty focusing given the switch to virtual instruction. In contrast, another student cited they felt they were able to meet the criteria as the process provided "leeway during this time of uncertainty." This comment was echoed in a subsequent response to which a student said they enjoyed that their voice was included in the process "especially [given the] uncertain time."

Relative to affective reactions to the intervention, students had mixed reviews. One, for example, said the process made them feel "neutral" and didn't feel "it made much of a difference" to their classroom experience. Another cited the experience as "a little uncomfortable." Overwhelmingly though, students felt positive about the process. One student said the process made them feel "Good!," while another appreciated "the instructors taking the time to allow us to build our own criteria and see what is working." That student went on to say: "it shows a lot of respect for our time and opinion."

One student posed that the process: "helped me understand the why behind some of the criteria, for example sometimes we are graded on things that don't seem important but being able to create what we are graded on makes sense."

When asked how the process affected student perceptions of the classroom and/or the instructor, students alleged feeling "listened to," which prompted a more positive outlook of the course because, "it showed that the instructor wanted to get our input." This led to perceptions of "really respecting" the instructor via students' perception that "they [the instructor] actually care." In fact, one student claimed they felt "closer to [their] TA than most of [their] TA's in the past." Finally, one student posed that the process: "helped me understand the why behind some of the criteria, for example sometimes we are graded on things that don't seem important but being able to create what we are graded on makes sense."

Finally, students responded to questions about what worked about the process and what could be improved. The majority of students felt the process worked as is, and particularly enjoyed breaking up into smaller groups in order to come up with criteria. This echoes findings that students appreciate the informality of small groups when wrestling with new ideas (Monsen et al., 2017). As mentioned earlier, however, a couple of students noted the extensiveness of the criteria as potentially overwhelming, with a lack of clarity as to which specific criterion to prioritize. This is consistent with research on classroom choice which cites marginal utility relative to the amount of choice offered (Patall et al., 2008). As a process, one student remarked, quite honestly, it was a "reminder to stay involved [but] feels a little like nagging though." With respect to improving the intervention, one student asked for a clearer understanding of how criteria mapped on to the grading scale, while another said they could have done with a shorter intervention on the first day of class.

Discussion

The current pilot study explored the effects of an intervention that sought to engage student voice in assessment (relative to participation evaluation) on their perceptions of power and attitudes towards assessment in a STEM classroom. Students reported a significant difference in their perceptions of power prior to and after implementation of the intervention, such that students felt stronger autonomy support in the classroom following the intervention. This follows findings in the realm of co-creation in curriculum where students express increased autonomy and motivation as a result of the process (Bergmark & Westman, 2015; Deeley & Bovill, 2017). A limitation to these findings includes an absence of control group which makes it difficult to isolate the intervention as the sole reason for differences found in student perceptions of power. Moreover, as students took the pre-survey on the first day of instruction their experience and ability to comment on that experience in the classroom, was limited. Nonetheless, these findings are encouraging considering the significant increase in student perception of their own power given the statistically small sample size and mere 10-week study period.

Overall, the intervention process appeared to bolster students' general perceptions of the classroom given their comments of feeling "listened to," "cared" about, and even, feeling "closer" to the instructor and the classroom.

In terms of the intervention experience, students largely appeared to find the process positive and helpful. While not a focus of the study, many students addressed the "uncertain times" that accompanied online instruction amidst a pandemic and cited the intervention as particularly useful therein. This has implications for virtual instructional methods, ways students can be involved in assessment dialogue despite not having a traditional learning environment, and how doing so may positively affect classroom perceptions.

This experience, however, was not intuitive for all participants. The discomfort noted from one student is perhaps a reminder of the traditional classroom perceptions of power where teachers are considered "the sole authority" in the classroom, and students "surrender" to that power as part of an "unwritten contract" (Sidky, 2017). By asking students to participate in the assessment realm, one they are not historically a part of, may lead to feelings of being unprepared and uncomfortable. This reinforces the idea of needing to engage student voice in assessment co-creation such that students become comfortable, and are ultimately prepared for a "lifetime of assessing their own learning" beyond the classroom (Boud & Falchikov, 2006, p. 400). The current intervention appears to begin this process, as cited by the student who said they now "understand the *why* behind criteria," as opposed to their previous experiences with grading and assessment. This comment is echoed in existing literature which suggests that involving students in the assessment process can provide an experience that "totally deepened my [student participant's] learning," by providing a metacognitive lens for what happens in the classroom and how these processes affect students' own learning (Cook-Sather, 2018, p. 927). While this intervention did not cover content-based assessment, it appeared to provide a small stepping-stone in getting students to understand the 'behind-the-scenes' of one aspect of assessment practice. Future iterations of this intervention should look to expand such work toward content-based assessments in the classroom.

Overall, the intervention process appeared to bolster students' general perceptions of the classroom given their comments of feeling "listened to," "cared" about, and even, feeling "closer" to the instructor and the classroom than they perhaps otherwise would have been (despite the virtual instructional setting). These findings of dialogue being advantageous towards students' perception of power in the classroom follow McLean's (2018) suggestion of explicit dialogue being essential to negotiate power and build trust between students and faculty. More generally, co-creation and partnership approaches in the classroom—validating and encouraging student voice and respective experience—begin the work of highlighting and subsequently tackling systems that lead to inequity (Cook-Sather, 2019). In fact, co-creation and the dialogue that accompanies it, has additionally been found to bolster the confidence of instructors towards explicitly addressing classroom inequities. Gauging faculty perceptions in future research of the co-creation process may be useful in validating such findings.

Future work using this intervention is currently underway to understand if the effects found in this pilot are replicable in traditional, in-person STEM classrooms, with larger sample sizes, and over longer periods of instructional time. More research in this area is necessary to understand how such an intervention affects perceptions across student identities, particularly

historically marginalized populations in higher education and STEM fields where the voices of such populations are not often represented in both the content and assessment of such content. Finally, future research should measure the effects of this intervention on other relevant outcome variables such as academic performance, feelings of belongingness, self-regulation, assessment anxiety, and retention/success in STEM.

Conclusion

The current pilot study aimed at engaging student voice in assessment practice, with the explicit purpose of studying student perceptions of power and attitudes towards assessment. While this study begins with involving students in non-content related assessment practice in a STEM classroom, the hope is to create an empirical foundation upon which research and practice can meaningfully incorporate student voice in content-related designs as well. More work in this area may help develop and validate this simple intervention for faculty to engage students in assessment without a complete overhaul of their existing assessment practices. Partnering with students at the assessment table may serve to empower and improve perceptions of the classroom, toward the end of fostering a more equitable learning environment for all students.

References

Bain, J. (2010). Integrating student voice: Assessment for empowerment. *Practitioner Research in Higher Education, 4*(1), 14-29.

Bean, J. C., & Peterson, D. (1998). Grading classroom participation. *New Directions for Teaching and Learning*, 1998(74), 33-40. https://doi.org/10.1002/tl.7403

Bergmark, U., & Westman, S. (2016). Co-creating curriculum in higher education: Promoting democratic values and a multidimensional view on learning. *International Journal for Academic Development, 21*(1), 28-40.

Black, A. E., & Deci, E. (2000). The effects of instructors' autonomy support and students' autonomous motivation on learning organic chemistry: A Self-Determination Theory perspective. *Science Education, 84*(6), 740-756.

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education, 31*(4), 399-413.

Bovill, C. (2020). Co-creation in learning and teaching: The case for a whole-class approach in higher education. *Higher Education, 79*(6), 1023-1037.

Brooman, S., Darwent, S., & Pimor, A. (2015). The student voice in higher education curriculum design: Is there value in listening? *Innovations in Education and Teaching International, 52*(6), 663-674. 910128

Bullen, H. A. (2012). Student-driven independent research projects: Developing a framework for success in analytical chemistry. *Analytical and Bioanalytical Chemistry, 404*(4), 927-930.

Chen, X., & Soldner, M. (2013). *STEM Attrition: College students' paths into and out of STEM fields* (p. 104). Washington, DC: National Center for Education Statistics.

Chirkov, V. I., & Ryan, R. M. (2001). Parent and Teacher Autonomy-Support in Russian and U.S. Adolescents: Common Effects on Well-Being and Academic Motivation. *Journal of Cross-Cultural Psychology, 32*(5), 618-635.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Hillsdale, N.J: L. Erlbaum Associates.

Committee on Barriers and Opportunities in Completing 2-Year and 4-Year STEM Degrees, Board on Science Education, Policy and Global Affairs, National Academy of Engineering, & National Academies of Sciences, Engineering, and Medicine. (2016). *Barriers and Opportunities for 2-Year and 4-Year STEM Degrees: Systemic Change to Support Students' Diverse Pathways*. (S. Malcom & M. Feder, Eds.). Washington, D.C.: National Academies

Cook-Sather, A. (2006). Sound, Presence, and Power: "Student Voice" in Educational Research and Reform. *Curriculum Inquiry, 36*(4), 359-390.

Cook-Sather, A. (2018). Listening to equity-seeking perspectives: how students' experiences of pedagogical partnership can inform wider discussions of student success. *Higher Education Research & Development, 37*(5), 923-936.

Cook-Sather, A. (2019). Respecting voices: How the co-creation of teaching and learning can support academic staff, underrepresented students, and equitable practices. *Higher Education*, 1-17.

Daily, S., Eugene, W., & Prewitt, D. (2007). The development of social capital in engineering education to improve student retention. In Proceedings of the ASEE Southeast Section Conference. Louisville, Kentucky.

Dancer, D., & Kamvounias, P. (2005). Student involvement in assessment: A project designed to assess class participation fairly and reliably. *Assessment & Evaluation in Higher Education, 30*(4), 445-454.

Deci, E., Connell, J. P., & Ryan, R. (1989). Self-determination in a work organization. *Journal of Applied Psychology, 74*(4).

Deci, E. L., & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology/Psychologie Canadienne, 49*(3), 182-185.

Deci, E., & Ryan, R. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum.

Deeley, S. J., & Bovill, C. (2017). Staff student partnership in assessment: enhancing assessment literacy through democratic practices. *Assessment & Evaluation in Higher Education, 42*(3), 463-477.

Deeley, S. J., & Brown, R. A. (2014). Learning through partnership in assessment. *Teaching and learning together in higher education*, *1*(13), 3.

Drewes, A., Scantlebury, K., & Soslau, E. (2019). Listening to the Missing Voices: Students' Perspectives on Coteachingdf. *Research & Practice in Assessment, 14*(Winter), 5-18.

Evans, M., & Boucher, A. R. (2015). Optimizing the Power of Choice: Supporting Student Autonomy to Foster Motivation and Engagement in Learning. *Mind, Brain, and Education, 9*(2), 87-91.

Fabert, N., Cabay, M., Rivers, M. B., Smith, M. L., & Bernstein, B. L. (2011). Exaggerating the typical and stereotyping the differences: Isolation experienced by women in STEM doctoral programs. *American Society for Engineering Education, 16*.

Falchikov, N. (2004). Involving students in assessment. *Psychology Learning and Teaching, 3*(2), 102-108.

Fletcher, A., & Shaw, G. (2012). How does student-directed assessment affect learning? Using assessment as a learning process. *International Journal of Multiple Research Approaches, 6*(3), 245-263.

French, R., Raven, B., & Cartwright, D. (1959). The bases of social power. *Classics of Organization Theory, 7*, 311-320.

Guberman, D. (2020). Grading together: Toward a partnership approach. *Teaching and Learning Together in Higher Education, 1*(29), 6.

Hewitt-Taylor, J. (2001). Self-directed learning: Views of teachers and students. *Journal of Advanced Nursing, 36*(4), 496-504.

Holec, H. (1981). *Autonomy and Foreign Language Learning*. Pergamon Press.

Martin-Hansen, L. (2018). Examining ways to meaningfully support students in STEM. *International Journal of STEM Education, 5*(1), 53.

McCroskey, J. C., & Richmond, V. P. (1983). Power in the classroom I: Teacher and student perceptions. *Communication Education, 32*(2), 175-184.

McGee, E. (2016). Devalued Black and Latino racial identities: A by-product of STEM college culture? *American Educational Research Journal, 53*(6), 1626-1662.

McLean, H. (2018). This is the way to teach: Insights from academics and students about assessment that supports learning. *Assessment & Evaluation in Higher Education, 43*(8), 1228-1240.

Meyer, K. R., & Hunt, S. K. (2011). Rethinking evaluation strategies for student participation. *Basic Communication Course Annual, 23*, 93-126.

Momsen, J. L., Long, T. M., Wyse, S. A., & Ebert-May, D. (2010). Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. *CBE—Life Sciences Education, 9*(4), 435-440.

Monsen, S., Cook, S., & Hannant, L. (2017). Students as partners in negotiated assessment in a teacher education course. *Teaching and Learning Together in Higher Education, 1*(21).

Pacharn, P., Bay, D., & Felton, S. (2013). The impact of a flexible assessment system on students' motivation, performance and attitude. *Accounting Education, 22*(2), 147-167.

Patall, E. A., Cooper, H., & Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: A meta-analysis of research findings. *Psychological Bulletin, 134*(2), 270-300.

Pretorius, L., van Mourik, G. P., & Barratt, C. (2017). Student choice and higher-order thinking: Using a novel flexible assessment regime combined with critical thinking activities to encourage the development of higher order thinking. *International Journal of Teaching and Learning in Higher Education, 29*(2), 389-401.

Rideout, C. A. (2018). Students' choices and achievement in large undergraduate classes using a novel flexible assessment approach. *Assessment & Evaluation in Higher Education, 43*(1), 68-78.

Rogers, S. L. (2013). Calling the question: Do college instructors actually grade participation? *College Teaching, 61*(1), 11-22.

Ryan, R. M., & Niemiec, C. P. (2009). Self-determination theory in schools of education: Can an empirically supported framework also be critical and liberating? *Theory and Research in Education, 7*(2), 263-272.

Ryan, R. M., & Weinstein, N. (2009). Undermining quality teaching and learning: A self-determination theory perspective on high-stakes testing. *Theory and Research in Education, 7*(2), 224-233.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119-144.

Sidky, G. (2017). The power game: Power dynamics between the teacher and the students in a graduate seminar. *English Language Teaching, 10*(5), 179.

Vandiver, D. M., & Walsh, J. A. (2010). Assessing autonomous learning in research methods courses: Implementing the student-driven research project. *Active Learning in Higher Education, 11*(1), 31-42.

van Griethuijsen, R. A., van Eijck, M. W., Haste, H., den Brok, P. J., Skinner, N. C., Mansour, N., ... & BouJaoude, S. (2015). Global patterns in students' views of science and interest in science. *Research in science education, 45*(4), 581-603.

Wanner, T., & Palmer, E. (2015). Personalising learning: Exploring student and teacher perceptions about flexible learning and assessment in a flipped university course. *Computers & Education, 88*, 354-369.

White, J. W. (2011). Resistance to classroom participation: Minority students, academic discourse, cultural conflicts, and issues of representation in whole class discussions. *Journal of Language, Identity & Education, 10*(4), 250-265.

Williams, G. C., & Deci, E. L. (1996). Internalization of biopsychosocial values by medical students: A test of Self-Determination Theory. *Journal of Personality and Social Psychology, 70*(4), 767-779.

Abstract

Using a critical paradigm, in this paper we highlight how current theoretical perspectives may serve to minimize and undermine historically Black colleges and universities (HBCUs) within the discourse on racial climate assessment in higher education. In particular, we closely examine a widely used campus climate theory to highlight how it centers predominantly White institutions and fails to consider the unique history, structure, and issues related to diversity, equity, and inclusion for HBCUs. In addition to identifying limitations on the current discourse on campus climate assessment at HBCUs, we provide important considerations and recommendations for future scholarship on this topic.

AUTHORS

Sosanya M. Jones, Ed.D.
Howard University

Glenn Allen Phillips, Ph.D.
Howard University

Re-Imagining Campus Climate Assessment at HBCUs

Efforts to assess campus climate in terms of inclusivity (not interpersonal violence) play a critical role in fulfilling student retention and institutional diversity, equity, and inclusion goals (Brennan, 2018; Cardemil, 2018; Museus et al., 2008). In the United States, interest in campus climate assessment has increased as institutions increasingly wrestle with their commitment to these principles (Steward, 2019). Today, a variety of assessment tools are available, and scholarship is beginning to examine their strengths, weaknesses, and overall value (Hurtado et al., 2008).

Problem Statement

Most of the discourse surrounding campus climate assessment has understandably revolved around predominantly White institutions (PWIs), which make up the majority of institutions of higher education in the United States. While this is an important and necessary focus, the approach leaves a troubling dearth of inquiry and reflection about the role of campus climate assessment in minority serving institutions (MSIs). In the United States, MSIs are federally designated institutions that serve a significant number of minoritized students. MSIs can be mission focused (i.e., designated to serve a particular minoritized group, such as African Americans, American Indians, or Hispanics) or enrollment focused (i.e., designated to serve a minimum percentage of enrolled minoritized students). There are seven types of MSIs in the United states, and each type has its own unique sociopolitical origin, political support and scrutiny, and target student population: Alaska Native or Native

CORRESPONDENCE

Email
sosanya.jones@howard.edu

Hawaiian serving institutions, Asian American or Native American Pacific Islander serving institutions, Hispanic serving institutions, historically Black colleges and universities, Native American serving Nontribal institutions, predominantly Black institutions, and tribal colleges or universities. This paper focuses on historically Black colleges and universities (HBCUs). As the oldest type of MSIs, HBCUs have a long history of serving disenfranchised students. Perhaps because of this history, there is often an implicit assumption that HBCUs do not have diversity, equity, and inclusion goals and requirements or do not need to engage in assessment in this area (Booker & Campbell-Whatley, 2019; Mutakabbir, 2018). The purpose of this paper is to raise critical questions and awareness about the need for more inquiry about campus climate assessment at HBCUs. We explicitly examine the limitations of the current discourse on campus climate assessment for HBCUs, but we believe this work highlights the value of similar work for other MSIs. We offer important considerations and recommendations for future scholarship on this topic.

Background

The term *campus climate* both enjoys and suffers from a broad definition in higher education. Its measurement or assessment is equally opaque. Rankin and Reason (2008) define *campus climate* as "attitudes, behaviors, and standards/practices that concern the access for, inclusion of, and level of respect for individual and group needs, abilities, and potential" (p. 264). Assessing campus climate, Hurtado, Carter, and Kardia (1998) argue, is "key for institutions that wish to create comfortable, diverse learning environments" (p. 53). Seen as a "proactive initiative rather than a reactive attempt to deal with significant issues affecting women, racial/ethnic minorities, disabled students, and LGBT students" (Hurtado et al., 2008, p. 204), campus climate assessment relies on a shared understanding of what and whose experience the particular assessment aims to understand.

The fluid definition of climate has complicated scholars' attempts to codify and create a reliable history of campus climate activities.

Hart and Fellabaum (2008)'s seminal analysis of 118 campus climate studies contributed to answering some of the field's earliest questions about campus climate research, including the following:

> What are the foci of the studies? Are they interested in race/ethnicity, gender, social class, or other issues? Is it the student experience that is most central, or is it faculty or staff, or a combination? What method or methods are being used? Who is conducting them? Are they being conducted by a campus researcher or an external researcher or consultant? (Hart & Fellabaum, 2008, p. 222)

It is difficult for researchers to do a thorough analysis of campus climate without first identifying what is meant by the term. However, the fluid definition of *climate* has complicated scholars' attempts to codify and create a reliable history of campus climate activities.

As a first step, Peterson and Spencer (1990) make the important distinction between campus culture and campus climate. Culture, they argue, "focuses on the deeply embedded patterns of organizational behavior and the shared values, assumptions, beliefs, or ideologies that members have about their organization or its work" (p. 6). Climate, on the other hand, "can be defined as the current common patterns of important dimensions of organizational life or its members' perceptions of and attitudes toward those dimensions" (p. 7). They further define *climate* by its major features:

- a primary emphasis on common participant views of a wide array of organizational phenomena that allow for comparison among groups or over time
- a focus on current patterns of beliefs and behaviors
- an often ephemeral or malleable character (Peterson & Spencer, 1990, p. 8)

Peterson and Spencer (1990) offer that "climate is pervasive, potentially inclusive of a broad array of organizational phenomena, yet easily focused to fit the researcher's or the administrator's interest" (p. 8). The "pervasive" and "inclusive" nature of climate opens the door to a variety of activities and assessments. However, the majority of available instruments

seem to target the experiences of underrepresented campus community members. Therefore, while the ways to identify climate are many, the targets for these assessments are narrower. While all students may be asked the degree to which they have experienced racism or micro-aggressions in the classroom, the data from students who are underrepresented are more likely to be reported and subsequently used for policy or programming initiatives (Jensen, 2011). This approach, however, can dangerously lead to othering the already othered and defining climate as what happens to others, not how the overrepresented are implicated in the experiences of others (Jensen, 2011).

Chang et al. (2010) explain that understanding climate is essential to helping colleges and universities meet diversity and equity related outcomes and indeed the student learning outcomes that form the core of their enterprise. However, the ambiguous parameters for the design of a more inclusive curriculum can lead to widely different standards for what is chosen and how more diverse materials are utilized. Chang et al. explain that "when it comes to engaging with diversity, White students tend to view this as an opportunity to be exposed to different cultures, whereas African American students tend to view this as an opportunity to enhance their institution's capacity for inclusion" (p. 46). This spectrum of purpose can lead many institutions to take an unbalanced approach to how they write campus climate assessments and make decisions based on the results. This can be even more complex for HBCUs that may be assumed to not have a problem with diversity and inclusion of historically marginalized populations.

Increasingly there have been specially tailored climate tools and models focused on specific populations, such as LGBT+ students (Evans et al., 2017; Garvey et al., 2015; Yost & Gilmore, 2011), international students (Soria & Brazelton, 2018), faculty (Austin, 1994; Martinez et al., 2014), socioeconomic class (Park et al., 2013), immigrant students (Stebleton et al., 2014), and various religious groups (Riggers-Piehl & Lehman, 2016). However, few of these studies identify HBCUs as research sites.

Paradigm

We approach our examination of this topic using a critical paradigm. The goal of critical theory is to uncover dominant perspectives that may serve to minimize, undermine, or devalue HBCUs within the discourse of assessment in higher education. Critical theory seeks to highlight hidden language, approaches, and perspectives that privilege and disempower in order to raise awareness, spark reflection, and invite actions that will promote change. In particular, we problematize the current discourse on diversity, equity, and inclusion assessment in higher education that centers PWIs and neglects to address HBCUs and the special circumstances, issues, and approaches that may need to be considered in order to implement effective assessment that measures how diversity, equity, and inclusion is implemented within these institutions.

The thrust of this paper is to ask what assessments of inclusion and belonging in terms of campus climate should look like in spaces where traditionally underrepresented populations are the majority.

Explanation of the Project

Recent campus climate research build on earlier iterations to include the experiences of racially and ethnically diverse students. Though distributed and analyzed with the stated purpose of making all students feel welcome and like they belong, these assessments are built on the embedded notion that other(ed) students do not belong and, in some cases, should not belong. The thrust of this paper is to ask what assessments of inclusion and belonging in terms of campus climate should look like in spaces where traditionally underrepresented populations are the majority. Before we explore the possibilities of how the discourse can be expanded to include HBCUs, it is important to understand the current discourse on campus climate assessment and its origins and influential factors.

Diversity, Equity, and Inclusion and the Courts

Higher education access for students of color has always been strongly contested by Whites, using the law of segregation as a shield. Once legal segregation was declared unconstitutional, historically White institutions resorted to using arbitrary methods to systematically exclude non-Whites (Goldstein Hode & Meisenbach, 2017). To combat this, affirmative action was established through executive orders. While regarded as a contentious

policy issue, desegregating acts like affirmative action contributed to institutionalizing diversity within higher education (Lipson, 2011). The overriding discourse of proponents of affirmative action points to the research on the benefits to the general (White) population (Goldstein Hode & Meisenbach, 2017). Some scholars have argued that this defense of affirmative action has negatively skewed the discourse on diversity, equity, and inclusion in higher education to center Whites as benefactors of these goals to the neglect of non-Whites, who get cast as educational props (Chin, 2011; Wray, 2008). Regardless, this discourse on affirmative action has bled into the way campus learning environments are conceptualized. The discourse focuses on how a critical mass of often racially diverse students can influence the learning environment for both White students, who stand to benefit from contact with diverse students, and students of color, who have been shown to do better in environments where they are at less risk for tokenism, alienation, and microaggressions (Campbell et al., 2019).

Access into higher education for students of color has always been strongly contested, thus limiting the opportunities for both Whites and students of color to learn from multiple perspectives, join in cross cultural dialogue, and bridge gaps between communities that have to interact in the work environment.

Campus Climate Theory and HBCUs

Hurtado et al.'s (2008) critical review of climate assessment instruments utilized a campus climate theory proposed by Hurtado, Clayton-Pederson, et al. (1998). In their review, they set current racial climate assessments against a four-dimensional backdrop of (a) historical, (b) structural, (c) psychological, and (d) behavioral dimensions of campus climate (Hurtado et al., 2008, p. 205). We have identified significant ways in which this framework may fall short for HBCUs.

Historical Component

As evidenced by the Hurtado et al. (2008) review of campus climate assessments, the historical component of campus climate theory is often ignored. The historical component of campus climate theory refers to the degree to which the institution has and sustains a "history and legacy of inclusion or exclusion" (Hurtado et al., 2008, p. 205). Indeed, history plays a significant role in campus climate. Efforts to address diversity, equity, and inclusion often focus on the exclusionary history of PWIs and efforts to eliminate this legacy of exclusion. Access into higher education for students of color has always been strongly contested, thus limiting the opportunities for both Whites and students of color to learn from multiple perspectives, join in cross cultural dialogue, and bridge gaps between communities that have to interact in the work environment. Universities have a long history of using arbitrary methods to decide who gets admitted while systematically using criteria that intentionally prohibit large numbers of students from particular groups from attending, such as SAT scores, pictures, and recommendation letters (Anderson, 1988; Bowen and Bok, 1998).

The discourse on campus climate and HBCUs is largely ahistorical (in that climate conversations do not include a rich contextualization of these colleges' and universities' beginnings). When history is acknowledged, there is often a failure to connect how diversity, equity, and inclusion may be complicated by this history. HBCUs are federally designated, defined, and protected according to the Higher Education Act of 1965 and arose in response to the theretofore unchecked traditions of discrimination and exclusion. Understanding their history, including the fact that HBCUs have historically been more open to diverse students than their PWI counterparts (Gasman & Nguyen, 2015), may help better contextualize efforts to assess their campus climate.

Structural Component

Although some efforts have been made to diversify higher education, there are still considerable disparities between the higher education enrollment and graduation rates of students of color and White students, especially at PWIs. Despite these gaps, a plethora of research demonstrates the benefits for White students of interacting with diverse, non-Black students at PWIs (Milem, 2003). However, sustained resistance to efforts to broaden access for students of color at PWIs persists.

With the settlement of several segregation de jure lawsuits between states and the Office of Civil Rights and the rise of performance-based funding, there has been mounting pressure for HBCUs to diversify and expand the type of students they recruit (Lundy-Wagner, 2015; Mobley et al., 2017). In fact, in some states, particular performance metrics designed

to meet state equity and diversity goals reward HBCUs for enrolling and supporting students who are characterized as racial minorities within the scope of the institution (i.e., non-Black students; Jones et al., 2017). To address this pressure, many HBCUs have begun to focus on recruiting more diversity by attracting and enrolling more non-Black, international, nontraditional, and LGBT+ student populations (Snipes & Darnell, 2017).

This new focus on diversification brings with it increased scrutiny and controversy about the identity and mission of HBCUs (Ingram et al., 2015; Palmer et al., 2018). While there is little discussion about these tensions when considering campus climate assessment, there needs to be more critique about how traditional campus climate assessments may gloss over these very real and complex tensions.

HBCUs, with their historical mission of serving disenfranchised and marginalized students, are closely connected to and greatly influenced by state government values and priorities and play a valuable role in promoting equality and opportunity in both education and the workforce. For example, HBCUs only represent 3% of all institutions of higher education in the United States but grant almost 20% of all bachelor's degrees earned by African Americans. They also produce 70% of all Black dentists and physicians, 50% of all Black engineers, 50% of all Black public school teachers, and 35% of all Black attorneys (Lomax, 2015). Additionally, many HBCUs have a history of and ongoing commitment to empowering both students of color and the communities from which they hail. This legacy is not often reflected in the ways in which HBCUs are assessed. We also know that enrollment among non-Black students at HBCUs is steadily growing. White students make up the highest number of non-Black enrollments, followed by Latinos and Asians respectively (Palmer, et.al., 2018).

HBCUs, with their historical mission of serving disenfranchised and marginalized students, are closely connected to and greatly influenced by state government values and priorities and play a valuable role in promoting equality and opportunity in both education and the workforce.

Psychological Component

Much of the literature on campus climate within higher education focuses on the psychological impact of the PWI campus climate for students of color as well the sociological implications for communities of color. Within PWIs, the campus environment for students of color can be particularly toxic, with stereotyping, tokenism, evidence of microaggressions, and overt racism increasing with the percentage of Whites in the student population (Harper & Hurtado, 2007; Karkouti, 2016). Empirical studies on PWIs show that inclusive programming positively engages students of color and provides a forum for promoting interactions between students of color and faculty members. This is obviously a crucial component in improving campus climate for students of color, who report feeling alienated and even unwelcome in a majority environment, but there is also evidence that White students benefit indirectly from this type of diversity through greater awareness of its presence and mission and directly by interacting with non-White peers. Consequently, White students seem to develop a higher level of empathy for diverse groups on campus (Bowman, 2010; Hurtado, 1999).

This kind of evidence is not apparent for non-African American and international students who attend HBCUs. For example, both Closson and Henry (2008) and Carter and Fountaine (2012) found that, absent a critical mass, White students who attend HBCUs do not feel isolated; however, they do experience feeling othered in certain contexts. The picture for Latinx students attending HBCUs shows they sometimes feel isolated and encounter microaggressions (Allen & Stone, 2016).

HBCUs embody inclusion in both their mission and history and for many Black students who attend HBCUs, the presence of a critical mass can make a significant impact on campus climate and feelings of self-efficacy and belonging. However, while HBCUs are producing better results for Black students, it is sometimes forgotten that a critical mass is not enough to cultivate inclusion or equity. Black students are not a monolith; there are subcultures within the Black student population, and every student has multiple identities that may also be marginalized. The danger in taking the success of Black student performance at HBCUs at face value is that it negates the intersectionality of Black students and deters a closer examination of how HBCUs may or may not nurture other parts of a student's identity and how they support non-Black students. For example, Bonner (2001) investigated HBCUs where Black women composed the highest percentage of students on campus and found that Black female students faced similar struggles within HBCU contexts as they did in PWIs,

reporting that HBCUs have a lot of work to do when it comes to dismantling sexism. These findings have been confirmed again and again by recent scholars (Glenn, 2019; Jean-Marie, 2017; Lockett & Gasman, 2018; Njoku & Patton, 2017). While HBCUs are more inclusive and supportive of diverse students in many ways, they still have work to do in creating a sense of belonging for non-Black student populations. For example, studies show that Muslim college students often encounter a Christian-normative environment at HBCUs, and while some HBCUs are making efforts to support non-Christian students, there is still a presumption and, in some cases, a demand to prioritize Christian beliefs and values. This norming of Christian values has implications for other populations as well, especially those in the LGBT+ community. In fact, according to Lenning (2017), "Historically Black Colleges and Universities (HBCUs) are notoriously perceived as unwelcoming towards lesbian, gay, bisexual, transgender, and queer (LGBT+) students, and are considerably behind predominately White institutions (PWIs) in regard to providing supportive and affirming environments" (p. 283). There is empirical evidence that LGBT+ students face a toxic environment at HBCUs, primarily due to the affiliation many HBCUs have to the Black church and its conservative fundamentalist interpretation of the Bible which condemns homosexuality (Coleman, 2016; Ward, 2005). Gasman and Nguyen (2015) also point out that the historical relationship between HBCUs and the church has stymied efforts to institutionalize support for LGBT+ students. Studies show that LGBT+ students at HBCUs often do not view their campus administrators and faculty as supportive (Lenning, 2017).

While HBCUs are more inclusive and supportive of diverse students in many ways, they still have work to do in creating a sense of belonging for non-Black student populations.

Behavioral Component

In PWIs, the behavioral component is addressed through change for greater inclusion. For example, White student development is enhanced through the efforts of students of color to organize protests against campus behavior that marginalizes students of color (Lane et al., 2017). In this setting, organized demonstrations empower students to become involved, raise awareness about identity issues formerly not known to the majority population, enhance democratic participation, and create a sense of purpose (Malaney & Berger, 2005). Astin (1993) found that cross-cultural discussions and interaction can increase racial understanding, foreign language skills, listening ability, and attendance at cultural events; while Kezar (2019) points to organized student groups and extracurricular activities that create spaces for learning, dialogue, and a shift in cultural inclusivity for different students. Samson (2018) presents evidence that largely homogenous White student organizations at PWIs, such as fraternities and sororities, have a negative impact on intercultural interactions for White students. Specifically, Samson suggests that there is a link between group membership in what he calls Greek letter organizations (GLOs) and heightened negative racial attitudes, particularly among White males. While these GLOs do not explicitly bar applicants or recruits from other races, implicit rules usually limit GLOs to token or trophy members of different races.

Another important area within higher education that addresses the behavior component is the classroom. Nussbaum (1997) refers to the social context of multiculturalism and the dilemma institutions face as they struggle with how to include diverse perspectives and ways of thinking into the curriculum. She asserts that all students benefit from a discourse of diversity that acknowledges and legitimizes marginalized histories, curriculums, and pedagogies.

There is an assumption that with a Black critical mass of students, HBCUs embrace behavioral inclusion. Perhaps this is why centralized offices and resources are not commonly dedicated to multicultural centers and diversity and equity at HBCUs (Carter & Christian, 2015). But while alternative curricula and attention to cultural identity are some of the key and most visible staples to an education at an HBCU, these are almost exclusively focused on the African American perspective. The key challenge is that the Black diaspora spans many continents, ethnicities, nationalities, classes, sexualities, and religions. There has been very little examination about the limits of what is considered Afrocentric or Black within HBCU curricula and approaches to teaching and learning. Additionally, the same challenges that were raised in regard to other marginalized identities and populations within the psychological area extend to the way space, traditions, artifacts, curriculum, pedagogy, and interactions are used and occur on the HBCU campus. For example, an HBCU with a

strong Christian doctrine may not consider how the celebration of other religious holidays may affect the academic calendar and availability of students from other faiths. Similarly, there may be pervasive heteronormative speech and language that conveys LGBT+ students are not welcome.

Denouement: Interrogating and Revising HBCU Climate Assessment

Moving forward, several steps can be taken to better conceptualize what an HBCU climate assessment could look like and how it could be used. First, Hurtado et al.'s (2008) seminal work must be extended or refocused to include the HBCU institutional family. Within this collection of schools, attention needs to be paid to how majority-minority HBCUs differ from majority-majority institutions in their approach to climate assessment. Looking across these institutions could provide necessary information on what is currently happening and the degree to which it is working.

Recommendations

While some scholars are beginning to push for the need for campus climate assessment at HBCUs (Contreras, 2018; Cuellar & Johnson-Ahorlu, 2016; Hurtado & Ruiz Alvarado, 2015), there is a noticeable dearth of approaches and models designed specifically for HBCUs. We hope that this conceptual work offers a foundation for HBCU climate assessment.

Proposal 1: A New Conceptual Theory for Campus Climate Tailored for HBCUs

Current campus climate assessments are deficient by nature as they operate based on the assumption that an alien or aberrant element is added to a monolithic (read functional) population. Climate assessment results are often focused on the responses of the minorities and rarely on the perceptions of the majority. In many ways, climate assessments at PWIs may be construed as a way to assess marginalized populations about their awareness of their place in institutions that were not built for them. New theories need to posit ways of knowing that transcend the them/us binary; identify intragroup and intergroup dynamics; accommodate the intersectional identities of all students; and suggest a definition of climate that goes beyond safety to include belonging, value, and ownership.

Climate assessments at PWIs may be construed as a way to assess marginalized populations about their awareness of their place in institutions that were not built for them. New theories need to posit ways of knowing that transcend the them/us binary.

In addition to reassessing traditional campus climate theories to consider the unique needs and strengths of HBCUs, we propose the consideration of other conceptual frameworks that decenter Whiteness and instead focus on the broader goal of equity. Two frameworks to consider are Gonzales et al.'s (2018) organizational framework and Bensimon et al.'s (2016) five guiding principles for addressing equity in policy and practice. Gonzales et al. (2018) propose using a new organizational framework for institutions to examine both individual and group dynamics. The goal of this framework is to not only improve the performance of the organization but also give more priority to the principles of power and justice. In this framework, the authors make an effort to reimagine a hierarchical construction of leadership so that leadership can be understood beyond de jure structure or what French and Raven (1959) refer to as *legitimate* power. Instead, referent power and expert power may better identify ways institutions manifest meaningful and sustainable transformation.

Bensimon et al.'s (2016) five guiding principles for equity in both policy and practice require the following: (a) clarity in language, goals, and measures; (b) equity-mindedness as a guiding paradigm for language and action; (c) equity in practice and policies designed to accommodate differences in the contexts of students' learning; (d) a continual process of learning, disaggregating data, and questioning assumptions about whether goals are relevant and effective; and (e) equity enacted as a system-wide principle. When applied to assessment, these five principles support a more in-depth, critical, and grounded approach to measuring diversity, equity, and inclusion in policy and practice. This theory also demands concrete language, objectives, and outcomes within assessments for supporting diversity and equity goals. While these principles are helpful for understanding how diversity, equity, and inclusion are addressed within policy, they have yet to be used as a framework for examining assessment. Though this proposal focuses specifically on the need for a new campus climate theory for HBCUs, the same need is certainly shared in their unique ways by the other MSIs.

Proposal 2: Creating New Campus Climate Assessments for HBCUs That Consider Their Unique Histories, Missions, Challenges, and Tensions

For HBCUs, this would require addressing specific subpopulations within the Black community and the intersectionality of Black students, LGBT+ students, religious minorities, students with disabilities, non-Black students of color, non-Black international students, and White students. HBCUs are not internally or externally monolithic. HBCUs have complex histories that tell an important narrative of American higher education. The many characteristics that can be used to define these institutions, their funding sources, resource richness, geography, longevity, and prestige can all influence how climate could (and should) be assessed. Internally, HBCUs must acknowledge that their student, faculty, and staff populations are not only diverse but that majority/minority politics cannot play out the same way at these institutions as they do at their predominantly White counterparts. Again, the spectrum of MSIs involves unique populations, histories, and contexts that should determine which assessments would be appropriate to capture their distinctive climate. Moreover, these new MSI-specific assessments must be accepted by accrediting bodies, state boards of education, professional organizations, and peer PWIs as equally valuable or internally more valuable than traditional assessments that may make more sense to those outside of the MSI world.

As accreditation can be an incredibly powerful force in institutional development and change, additional standards or requirements related to climate could motivate institutions to regularly assess diversity, equity, and inclusion and think about how these areas are linked to other campus-wide goals.

Proposal 3: Linking Climate Assessment to Accreditation

Accreditation is a critical process designed to foster continuous improvement and the development of exemplary programs. In the accountability movement, accreditation has become more visible and significant to an institution's survival. While primarily loss of accreditation translates to a loss in Title IV funding, it can also be tied to a loss in research dollars, enrollment, alumni giving, and prestige. All colleges and universities have the goal of achieving and maintaining accreditation for all of their academic units; however, HBCUs are more vulnerable during the accreditation process. A socially just and inclusive mission often leads HBCUs to accept more underprepared students, which can have a direct impact on student learning outcome assessments and graduation outcomes.

With the decline of state funding and the rise of accountability demands, particularly those related to regional and discipline-specific accreditation, assessment has become a top priority in higher education. Most institutions of higher education engage in some form of assessment, but the push to identify and incorporate ways of assessing diversity, equity, and inclusion is rarely a part of most institutions' accreditation-facing assessment strategy. Assessment for continuous improvement usually focuses on learning, graduation outcomes, financial efficiency, and mission-centered effectiveness. As accreditation can be an incredibly powerful force in institutional development and change, additional standards or requirements related to climate could motivate institutions to regularly assess diversity, equity, and inclusion and think about how these areas are linked to other campus-wide goals. Accreditation agencies rarely ask questions about diversity, equity, and inclusion unless they are specifically noted in the strategic plan or mission of an institution, but many regional accreditors will require institutions to disaggregate their student and faculty outcomes by a variety of demographics, including race, gender, expected family contribution, and faculty employment status. An important step forward would be for these agencies to include race and ethnicity, as appropriate, when assessing equity.

Campus climate assessment provides an important but additional accountability lever that is usually only pulled by institutions whose mission specifically identifies a focus on inclusion or diversity. Little consideration is given to how HBCUs cultivate diversity and equity goals. There is a growing critique about the lack of attention paid to these values in the accreditation process, and how addressing diversity and equity can benefit HBCUs in the accreditation process. Additional regional requirements and discipline-specific requirements could force institutions to acknowledge that climate assessment extends beyond tracking and means more than good publicity.

Proposal 4: Rethink How Campus Climate Assessment Focuses on Outcomes

Cardemil (2018) points out that while campus climate assessment can play a critical role in advancing diversity and inclusion efforts, one of the key limitations of this type of assessment is that it focuses on outcomes, not processes. He suggests that the approach to campus climate assessment needs to change to a more developmental approach that is reflective and educationally process-centered rather than focused on outcomes. Whether building on Cardemil's approach or utilizing more traditional assessments, outcomes and the use of data for improvement must be reconsidered. Institutions cannot depend on head counts and data from the multicultural center to give the kind of 360 degree view of climate that is needed for administration to make lasting and appropriately funded changes.

It is time to reimagine campus climate theory and assessment to build and create sustainable theory and practical campus climate assessment models for the 21st century that emphasize the enduring mission and goal of HBCUs to offer equitable, diverse, and inclusive learning environments for all students.

Scholarly Significance and Suggestions for Future Research

In both accreditation and assessment efforts, diversity, equity, and inclusion at HBCUs are woefully under researched. This paper offers a bridge for assessment officers to consider the approaches, tools, and gaps for assessing institutional commitment and support toward these values and goals. In addition to the aforementioned recommendations, we have identified possibilities and openings for future research that may be of interest to scholars interested in campus climate assessment at HBCUs and other MSIs.

Conclusion

Many HBCUs are currently reevaluating their institutional strategic plan and assessments due to demands from the state, federal government, and accreditation organizations. HBCUs are already doing a fantastic job of educating underrepresented students of color in the United States (Chenier & Bista, 2019; Palmer et al., 2010; Toldson, 2018). In many areas they have done a better job than PWIs. Adopting specially tailored campus climate assessment would not only inform how these institutions can best meet the needs of their different populations of students but also highlight their achievements. Traditional models of campus climate assessment may not be suitable for HBCUs because they are based on theoretical suppositions that center PWIs and White students and do not take into account the unique history, structure, populations, and tensions found within HBCUs. It is time to reimagine campus climate theory and assessment to build and create sustainable theory and practical campus climate assessment models for the 21st century that emphasize the enduring mission and goal of HBCUs to offer equitable, diverse, and inclusive learning environments for all students.

AUTHOR'S NOTE

Portions of these findings were presented as a research presentation at the 2019 Association for the Study of Higher Education (ASHE) conference. We have no known conflict of interest to disclose.

References

Allen, T. O., & Stone, A. N. (2016). Degrees of separation: Latino students' transitions to a Texas HBCU. *Journal of Student Affairs Research and Practice, 53*(4), 378-390.

Anderson, J. (1988). *The education of Blacks in the South, 1860-1935*. University of North Carolina Press.

Astin, A. W. (1993). Diversity and multiculturalism on campus: How are students affected? *Change: The Magazine of Higher Learning, 25*(2), 44-49.

Austin, A. E. (1994). Understanding and assessing faculty cultures and climates. *New Directions for Institutional Research, 1994*(84), 47-63.

Bensimon, E. M., Dowd, A. C., & Witham, K. (2016). *Five principles for enacting equity by design*. Association of American Colleges & Universities.

Bonner, F. B. (2001). Addressing gender issues in the historically Black college and university community: A challenge and call to action. *Journal of Negro Education, 70*(3), 176-191. https://www.jstor.org/stable/3211209

Booker, K. C., & Campbell-Whatley, G. (2019). Student perceptions of inclusion at a historically Black university. The *Journal of Negro Education, 88*(2), 146-158. https://www.jstor.org/stable/10.7709/jnegroeducation.88.2.0146

Bowen, W., & Bok, D. (1998). *The shape of the river: Long-term Consequences of Considering Race in College and University Admissions*. Princeton University Press.

Bowman, N. A. (2010b). Disequilibrium and resolution: The nonlinear effects of diversity courses on well-being and orientations toward diversity. *The Review of Higher Education, 33*(4), 543-568. https://psycnet.apa.org/doi/10.1353/rhe.0.0172

Brennan, M. (2018). *Diversity metrics, measurement, and evaluation*. http://www.workforcediversitynetwork.com/res_articles_DiversityMetricsMeasurementEvaluation.asp

Campbell, S. D., Carter-Sowell, A. R., & Battle, J. S. (2019). Campus climate comparisons in academic pursuits: How race still matters for African American college students. *Group Processes & Intergroup Relations, 22*(3), 390-402. https://doi.org/10.1177%2F1368430218823065

Cardemil, E. (2018). The gap between assessment and creating inclusion. In S. Gertz, B. Huang, & L. Cyr (Eds.), *Diversity and inclusion in higher education and societal contexts* (pp. 271-284). Palgrave Macmillan.

Carter, J. D., & Christian, W. A., III. (2015). An analysis of diversity work at HBCUs and PWIs. In T. N. Ingram, D. Greenfield, J. D. Carter, & A. A. Hilton (Eds.), *Exploring issues of diversity within HBCUs* (pp. 107-134). Information Age Publishing.

Carter, J. D., & Fountaine, T. P. (2012). An analysis of White student engagement at public HBCUs. *Educational Foundations, 26*, 49-66. https://files.eric.ed.gov/fulltext/EJ1000230.pdf

Chang, M. J., Milem, J. F., & Antonio, A. L. (2010). Campus climate and diversity. In J. H. Schuh, S. R. Jones, & S. R. Harper (Eds.), *Student services: A handbook for the profession* (pp. 43-58). John Wiley & Sons.

Chenier, F., & Bista, K. (2019). What works at Historically Black Colleges and Universities (HBCUs). *Journal of Underrepresented & Minority Progress, 3*(1), 97-99.

Chin, J. (2011). What a load of hope: The post-racial mixtape. *California Western Law Review, 48*(2), 369. https://scholarlycommons.law.cwsl.edu/cwlr/vol48/iss2/8

Steward, P. (2019, October 17). Campus climate survey findings break along color line. *Diverse Issues in Higher Education*. https://diverseeducation.com/article/157754/October 17, 2019.

Closson, R. B., & Henry, W. J. (2008). The social adjustment of undergraduate White students in the minority on an historically Black college campus. *Journal of College Student Development, 49*(6), 517-534. https://psycnet.apa.org/doi/10.1353/csd.0.0036

Coleman, K. (2016). The Difference Safe Spaces Make: The Obstacles and Rewards of Fostering Support for the LGBT Community at HBCUs. *SAGE Open*. https://doi.org/10.1177/2158244016647423

Contreras, F. (2018). Latino faculty in Hispanic-serving institutions: Where is the diversity? *Association of Mexican American Educators Journal, 11*(3), 223-250. https://doi.org/10.24974/amae.11.3.368

Cuellar, M., & Johnson-Ahorlu, R. N. (2016). Examining the complexity of the campus racial climate at a Hispanic serving community college. *Community College Review, 44*(2), 135-152. https://doi.org/10.1177%2F0091552116632584

Evans, R., Nagoshi, J. L., Nagoshi, C., Wheeler, J., & Henderson, J. (2017). Voices from the stories untold: Lesbian, gay, bisexual, trans, and queer college students' experiences with campus climate. *Journal of Gay & Lesbian Social Services, 29*(4), 426-444.

French, J. R. P., Jr., & Raven, B. H. (1959). The bases of social power. In D. Cartwright (Ed.), *Studies in social power* (pp. 150-167). Institute for Social Research.

Garvey, J. C., Taylor, J. L., & Rankin, S. (2015). An examination of campus climate for LGBTQ community college students. *Community College Journal of Research and Practice, 39*(6), 527-541. https://doi.org/10.1080/10668926.2013.861374

Gasman, M., & Nguyen, T.-H. (2015). Myths dispelled: A historical account of diversity and inclusion at HBCUs. *New Directions for Higher Education, 2015*(170), 5-15. https://doi.org/10.1002/he.20128

Glenn, C. (2019, April). An intersectional analysis of African-American females at Historically Black Colleges and Universities. In *ICGR 2019 2nd International Conference on Gender Research* (p. 298). Academic Conferences and publishing limited.

Goldstein Hode, M., & Meisenbach, R. J. (2017). Reproducing whiteness through diversity: A critical discourse analysis of the pro-affirmative action amicus briefs in the Fisher case. *Journal of Diversity in Higher Education, 10*(2), 162.

Gonzales, L. D., Kanhai, D., & Hall, K. (2018). Reimagining organizational theory for the critical study of higher education. In M. B. Paulsen (Ed.), *Higher education: Handbook of theory and research* (Vol. 33, pp. 505-559). Springer International Publishing.

Harper, S. R., & Hurtado, S. (2007). Nine themes in campus racial climates and implications for institutional transformation. *New Directions for Student Services, 2007*(120), 7-24. https://doi.org/10.1002/ss.254

Hart, J., & Fellabaum, J. (2008). Analyzing campus climate studies: Seeking to define and understand. *Journal of Diversity in Higher Education, 1*(4), 222-234. https://psycnet.apa.org/doi/10.1037/a0013627

Hurtado, S. (1999). Reaffirming educators' judgment: Educational value of diversity. *Liberal Education, 85*(2), 24-31.

Hurtado, S., Carter, D. F., & Kardia, D. (1998). The climate for diversity: Key issues for institutional self study. *New Directions for Institutional Research, 1998*(98), 53-63. https://doi.org/10.1002/ir.9804

Hurtado, S., Clayton-Pedersen, A. R., Allen, W. R., & Milem, J. F. (1998). Enhancing campus climates for racial/ethnic diversity: Educational policy and practice. *The Review of Higher Education, 21*(3), 279-302. doi:10.1353/rhe.1998.0003

Hurtado, S., Griffin, K., Arellano, L., & Cuellar, M. (2008). Assessing the value of climate assessments: Progress and future directions. *Journal of Diversity in Higher Education, 1*(4), 204-221.

Hurtado, S., & Ruiz Alvarado, A. (2015). *Discrimination and bias, underrepresentation, and sense of belonging on campus*. https://vtechworks.lib.vt.edu/handle/10919/83064

Ingram, T. N., Greenfield, D. F., Carter, J. D., & Hilton, A. A. (2015). Examining the significance of HBCUs and their inclusive climate. In T. N. Ingram, D. F. Greenfield, J. D. Carter, & A. A. Hilton (Eds.), *Exploring issues of diversity within HBCUs* (pp. 1-4). Information Age Publishing.

Jean-Marie, G. (2017). Black women at the helm in HBCUs: Paradox of gender and leadership. In *Black colleges across the diaspora: Global perspectives on race and stratification in postsecondary education*. Emerald Publishing Limited.

Jensen, S. Q. (2011). Othering, identity formation and agency. *Qualitative Studies, 2*(2), 63-78.

Jewell, J. O. (2002). To set an example: The tradition of diversity at historically Black colleges and universities. *Urban education, 37*(1), 7-21. https://doi.org/10.1177%2F0042085902371002

Jones, T., Jones, S., Elliott, K. C., Russell Owens, L., Assalone, A. E., & Gándara, D. (2017). *Outcomes based funding and race in higher education*. Springer.

Karkouti, I. M. (2016). Professional leadership practices and diversity issues in the US higher education system: A research synthesis. *Education, 136*(4), 405-412. https://eric.ed.gov/?id=EJ1104195

Kezar, A. (2019). *Creating a diverse student success infrastructure: The key to catalyzing cultural change for today's student*. Pullias Center for Higher Education.

Krumer-Nevo, M., & Sidi, M. (2012). Writing against othering. *Qualitative inquiry, 18*(4), 299-309. https://doi.org/10.1177%2F1077800411433546

Lane, T. Y., Chiarelli-Helminiak, C., Bohrman, C., & Lewis, T. (2017). The teachable moment: Engaging students in social justice movements. *Social Work Education, 36*(4), 359-372. https://doi.org/10.1080/02615479.2016.1268120

Lee, J. M., Jr., & Keys, S. W. (2013). *Land-grant but unequal: State one-to-one match funding for 1890 land-grant universities* (Report No. 3000-PB1). Association of Public and Land-Grant Universities.

Lenning, E. (2017). Unapologetically queer in unapologetically Black spaces: Creating an inclusive HBCU campus. *Humboldt Journal of Social Relations, 1*(39), 283-293. https://digitalcommons.humboldt.edu/cgi/viewcontent.cgi?article=1019&context=hjsr

Lipson, D. N. (2011). The resilience of affirmative action in the 1980s: Innovation, isomorphism, and institutionalization in university admissions. *Political Research Quarterly, 64*(1), 132-144. https://doi.org/10.1177%2F1065912909346737

Lockett, A. W., & Gasman, M. (2018). Envisioning equity: Women at the helm of HBCU leadership', underserved populations at Historically Black Colleges and Universities *Diversity in Higher Education*, Volume 21. https://doi.org/10.1108/S1479-364420180000021014

Lomax, M. (2015). *The condition of college & career readiness 2014*. United Negro College Fund. https://cdn.uncf.org/wpcontent/uploads/reports/2014_The_Condition_of_College_Career_Readiness_2014_AfAm_ACT.pdf

Lundy-Wagner, V. C. (2015). Coming out of the shadows: Rethinking the education policy agenda for diversity and HBCUs. *New Directions for Higher Education, 2015*(170), 91-101. https://doi.org/10.1002/he.20134

Malaney, G. D., & Berger, J. B. (2005). Assessing how diversity affects students' interest in social change. *Journal of College Student Retention: Research, Theory & Practice, 6*(4), 443-460. https://scholarworks.umass.edu/cie_faculty_pubs/8

Martinez, V., Miller, M. H., & Tyson, W. (2014). Analyzing STEM faculty demographics and faculty climate survey. In P. J. Gilmer, B. Tansel, & M. H. Miller (Eds.), *Alliances for advancing academic women* (pp. 79-96). SensePublishers. https://doi.org/10.1007/978-94-6209-604-2_4

Milem, J. F. (2003). The educational benefits of diversity: Evidence from multiple sectors. In M. J. Chang, D. Witt, J. Jones, & K. Hakuta (Eds.), *Compelling interest: Examining the evidence on racial dynamics in higher education* (pp. 126-169). Stanford University Press.

Mobley, S. D., Jr., Daoud, N., & Griffin, K. A. (2017). Re-coloring campus: Complicating the discourse about race and ethnicity at historically Black colleges and universities. In M. C. Brown & T. E. Dancy (Eds.), *Black colleges across the diaspora: Global perspectives on race and stratification in postsecondary education* (Vol. 14, pp. 29-47). Emerald Publishing Limited.

Museus, S. D., Nichols, A. H., & Lambert, A. D. (2008). Racial differences in the effects of campus racial climate on degree completion: A structural equation model. *The Review of Higher Education, 32*(1), 107-134. https://psycnet.apa.org/doi/10.1353/rhe.0.0030

Mutakabbir, Y. T. (2018). Religious minority students at HBCUs. In C. H. Davis, A. A. Hilton, & D. L. Outten (Eds.), *Underserved populations at historically Black colleges and universities: The pathway to diversity, equity, and inclusion* (pp. 107-120). Emerald Publishing Limited. https://doi.org/10.1108/S1479-364420180000021019

Njoku, N. R., & Patton, L. D. (2017). Explorations of respectability and resistance in constructions of Black womanhood at HBCUs. *Critical perspectives on Black women and college success*, 143-157.

Nussbaum, M. (1997). Cultivating humanity: A classical defense of reform in liberal education. Harvard University Press.

Palmer, R. T., Arroyo, A. T., & Maramba, D. C. (2018). Exploring the perceptions of HBCU student affairs practitioners toward the racial diversification of Black colleges. *Journal of Diversity in Higher Education, 11*(1), 1-15. https://psycnet.apa.org/doi/10.1037/dhe0000024

Palmer, R. T., Davis, R. J., & Maramba, D. C. (2010). Role of an HBCU in supporting academic success for underprepared Black males. *Negro Educational Review, 61*, 85-106. https://www.questia.com/library/journal/1P3-2216432291/role-of-an-hbcu-in-supporting-academic-success-for

Park, J. J., Denson, N., & Bowman, N. A. (2013). Does socioeconomic diversity make a difference? Examining the effects of racial and socioeconomic diversity on the campus climate for diversity. *American Educational Research Journal, 50*(3), 466-496. https://doi.org/10.3102%2F0002831212468290

Peterson, M. W., & Spencer, M. G. (1990). Understanding academic culture and climate. *New Directions for Institutional Research, 1990*(68), 3-18. https://doi.org/10.1002/ir.37019906803

Rankin, S., & Reason, R. (2008). Transformational tapestry model: A comprehensive approach to transforming campus climate. *Journal of Diversity in Higher Education, 1*(4), 262-274. https://psycnet.apa.org/doi/10.1037/a0014018

Riggers-Piehl, T. A., & Lehman, K. J. (2016). Modeling the relationship between campus spiritual climate and the sense of belonging for Christian, Muslim, and Jewish students. *Religion & Education, 43*(3), 247-270.

Samson, F. L. (2018). *Fraternity membership and negative racial attitudes among U.S. college students*. Institute for Research on Labor and Employment at UCLA. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3187953

Snipes, J. T., & Darnell, C. (2017). Non-Black student recruitment at historical Black colleges and universities. In M. C. Brown & T. E. Dancy (Eds.), *Black colleges across the diaspora: Global perspectives on race and stratification in postsecondary education* (Vol. 14, pp. 49-67). Emerald Publishing Limited. https://doi.org/10.1108/S1479-358X20160000014003

Soria, K. M., & Brazelton, G. B. (2018). International students' experiences with campus climate at large, public research universities. In K. M. Soria (Ed.), *Evaluating campus climate at US research universities* (pp. 251-275). Palgrave Macmillan.

Stebleton, M. J., Soria, K. M., & Huesman, R. L. (2014). Recent immigrant students at research universities: The relationship between campus climate and sense of belonging. *Journal of College Student Development, 55*(2), 196-202. https://doi.org/10.1353/csd.2014.0019

Toldson, I. A. (2018). Why Historically Black Colleges and Universities are Successful with Graduating Black Baccalaureate Students Who Subsequently Earn Doctorates in STEM (Editor's Commentary). *The Journal of Negro Education, 87*(2), 95-98.

Ward, E. G. (2005). Homophobia, hypermasculinity and the US Black church. *Culture, Health & Sexuality, 7*(5), 493-504. https://doi.org/10.1080/13691050500151248

Wray, A. (2008). The rhetoric of racism: First-year composition students on affirmative action. *Watermark Journal, 2*, 184-201. https://web.csulb.edu/journals/watermark/pdf/Watermark%20vol2%202008.pdf

Yost, M. R., & Gilmore, S. (2011). Assessing LGBTQ campus climate and creating change. *Journal of Homosexuality, 58*(9), 1330-1354. https://doi.org/10.1080/00918369.2011.605744

Abstract

Writing support programs for students in writing-intensive, disciplinary courses are well established and take many forms, including communication centers, web-based skill development programs, and embedded writing consultants. This paper assesses the effectiveness of a program that embeds a writing grader, who assesses only the grammar of students' submissions, to encourage and support business faculty in including written assignments. Our analysis of grammar errors across three writing assignments showed that students rarely included Status Marking Errors (e.g., nonstandard verb forms and double negatives) and did reduce errors from the first to last assignments. However, the cause of the error reduction and the program's long-term effectiveness in improving students' grammar skills is inconclusive. Based on our findings, we offer recommendations to program organizers for better aligning a program's stated and practical goals when providing writing support.

AUTHORS

Lindsay C. Clark, Ph.D.
Sam Houston State University

Zijun Luo, Ph.D.
Sam Houston State University

Ashly Bender Smith, Ph.D.
Sam Houston State University

Effects of Course-Embedded Grammar Graders: Evidence from a Business College Writing Initiative Program

Initiatives to support and develop student writing skills have a long history in composition, writing studies, writing center, and writing across the curriculum (WAC) research and pedagogy. These support programs may function at the university level through writing centers, working with students one-on-one to improve their writing skills, and through WAC programs that help faculty integrate writing into their courses. In discipline-based courses, other student-focused services like writing fellows or embedded tutors or consultants serve to bridge the gap between writing assistance and writing instruction (Carpenter et al., 2014). Specifically, the role of this embedded support is to guide students through course-specific writing assignments by engaging students in the writing process and providing feedback.

Feedback on student writing is widely discussed across many higher education disciplines, with research on the scope, format, and type of feedback most impactful for strengthening students' writing skills. To encourage students to engage in the writing process, many researchers have studied the potential for alternative feedback formats, including written versus digital comments (Grouling, 2018) and written versus audio feedback (Keane et al., 2018). Additional studies have explored how corrective and formative feedback might incentivize or motivate students to engage in revision processes, as well as facilitate communication skill development (Bitchenor & Knoch, 2010; Yu et al., 2020).

CORRESPONDENCE

Email
lclark@shsu.edu

Best practices for improving students' communication skills continue to be an important discussion in higher education, as many studies have reported both increased value for, as well as deficiencies in, the writing ability of college students entering the workforce [Addams & Allred, 2015; National Association of Colleges and Employers (NACE), 2018]. These findings place particular pressure on colleges of business aiming to prepare students to be effective communicators in various professional industries. In response, many business schools have integrated a variety of writing support interventions, including in-house communication and writing centers (Caldwell & Al-Ajmi, 2018), web-based writing skills programs (Austin et al., 2018), and grammar and mechanics instructional strategies (O'Neill, 2018; Quible, 2006; Willis et al., 2012).

Additional research on written communication skills reiterates the importance of administering a writing support initiative focused on assessing common grammar errors to focus feedback, revision, instruction, and assessment strategies.

Still, grammar and mechanics instruction remains a highly debated and frequently identified area of improvement for college students entering the workforce. Willis et al. (2012) administered a survey to over 600 business undergraduates, determining that students' abilities to identify and correct common grammar and mechanical errors were severely lacking. Several studies have demonstrated how simply telling students how many errors they made without providing formative feedback actually results in increased writing anxiety, especially if students previously had low self-efficacy about their writing (Ekholm et al., 2015; Mascle, 2013; Zumbrunn et al., 2016). Similarly, O'Neill (2018) assessed which grammar errors her business students made most frequently, arguing that most students make errors in only a few concentrated areas of punctuation and style. She argued that with "critical and analytical thinking, but without mechanics to ensure conciseness and clarity, writers can miss the opportunity to bring their ideas to wider audiences" (p. 9). Additional research on written communication skills reiterates the importance of administering a writing support initiative focused on assessing common grammar errors to focus feedback, revision, instruction, and assessment strategies.

In the authors' college of business, concern regarding business majors' grammar knowledge led a business communication faculty member to create a resource, a Credibility Killers handout, which describes ten grammar errors found to be most noticed by a business professional (see Appendix A). The included errors were developed based on research in Business Communication and Rhetoric and Composition (e.g., Beason, 2001; Gray & Heuser, 2003; Hairston, 1981; Lunsford & Lunsford, 2008; O'Neill, 2018; Sigmar & Austin, 2015). The Credibility Killers handout includes two categories of errors: Status Marking Errors and Serious Errors. The Status Marking Errors include a) nonstandard verb forms, b) lack of verb-subject agreement, c) double negatives, and d) object pronoun as subject. The Serious Errors include e) sentence fragments; f) run-on sentences; g) non-capitalization of proper nouns; h) misspelled words; and i) comma errors, such as clause and comma series errors. These categories and included error types seem to have been first delineated in Maxine Hairston's 1981 article, "Not All Errors are Created Equal: Nonacademic Readers in the Professions Respond to Lapses in Usage," and researchers have continued to test the effect of grammar errors using Hairston's categories (e.g., Gray & Heuser, 2003).

Developed from a desire to support students' writing development in upper-division business courses, the Credibility Killers handout was circulated across the college and soon after was used as a "rubric" for a new writing support initiative focused on helping students identify and correct grammar errors in their writing. This writing support program, named the Writing Initiative, is the focus of this study[1]. As part of the business college's communication assessment committee, we offer preliminary findings regarding if and how this initiative influences students' written communication skills, specifically the correction of grammar and mechanical errors.

The Writing Initiative Overview

The Writing Initiative (WI) at the authors' university is intended to support students' writing skill development by encouraging faculty to incorporate writing assignments into their courses. The program was developed in response to communication assessment results

[1] The authors thank Dr. Christopher Cassidy, Jennifer Ormond, the involved student-workers, and the Communication Goals Assessment Team for their assistance in collecting and coding the Writing Initiative assessment samples.

the WI's effect on participating students' performance had not been completed. Therefore, we investigated if and how the WI might influence students' grammatical correctness, and herein we report the results of our assessment and discuss implications for both future assessment and classroom instruction. The results revealed opportunities for improvement of our writing support initiative. We conclude with suggestions on ways organizers can improve writing support programs to ensure alignment between support goals and practices.

Method

During the assessment semester, 10-12 faculty participated in the WI but only two used the program for multiple assignments, allowing the assessment of error trends. One of these two faculty members was teaching an Executive Master of Business Administration (EMBA) course with 10 enrolled students. The other faculty member was teaching two sections of the senior-level capstone course required of all business majors, Strategic Management and Policy. The professor assigned four independent writing assignments throughout the semester that were submitted to the WI. Each assignment required a written submission of approximately 1,000 words from the student. Fifty-three students submitted a paper for one or more of the four assignments, with a total of 174 graded submissions. Overall, the Strategic Management and Policy course provided an opportunity to identify the potential influence of the WI across multiple assignments submitted by a selection of students from across the business major who were at or near the end of their program.

Overall, the Strategic Management and Policy course provided an opportunity to identify the potential influence of the WI across multiple assignments submitted by a selection of students from across the business majors who were at or near the end of their program.

The WI grader tallied the majors errors in all submissions for the four assignments. For each assignment, each student's errors were tallied in an Excel file. In addition to recording the total error count for each student, the number of each type of error was recorded. Only the errors from the Credibility Killers list, the "major errors," were counted for the assessment, though the grader did provide additional feedback to the students. Only one WI grader evaluated the assignments submitted in the two sections included in this assessment.

The statistical analysis was carried out at two levels. We first analyzed aggregate trends with the full sample. For each assignment, we computed the mean, standard deviation, maximum, and the percentage of students with no or only one error. We also conducted a comparative study between our findings and those in Lunsford and Lunsford (2008). We then looked at the 30 students who submitted all four assignments. This individual analysis allowed us to investigate the number of students that have improved throughout the semester and to parse errors by assignment. Detailed results of this statistical analysis are presented in the next section.

Results

The Credibility Killers rubric categorizes writing errors into nine areas (A-I). However, hardly any students made mistakes in areas A through D, the Status Marking Errors. As a result, this analysis focuses on categories E through I, the Serious Errors: e) sentence fragments, f) run-on sentences, g) non-capitalization of proper nouns, h) misspelled words, and i) comma errors.

The following analyses focus on Assignments 1, 3, and 4. Due to the structure of the assignments, Assignment 2 showed peculiar behavior and hence has been excluded from the analysis. In addition to analyses of the full sample of 174 assignments from 53 students, a sub-set of the sample is analyzed. The sub-set includes a group of 30 students who submitted to all four assignments.

Aggregate Analysis (Full Sample)

Table 1 shows the mean, standard deviation, and maximum of the five Serious Errors categories for Assignments 1, 3, and 4, from the full sample. The first panel shows that the average number of errors that occurred per assignment in each category declined between the first and fourth assignments. From Assignment 1 to 4, there was over 30% reduction in errors in all categories, with the smallest change in category I (Comma Errors) at 32.76% and the largest change in category G (Non-capitalization) at 70.59%.

Table 1
Summary Statistics for Assignments 1, 3, & 4

	A1	A3	A4	% Reduction from A1 to A4
Mean				
E (Sentence Fragments)	.76	.49	.43	43.42%
F (Run-on Sentences)	.78	.56	.49	37.18%
G (Non-capitalization)	.51	.24	.15	70.59%
H (Misspelled Words)	3.07	2.84	1.87	39.09%
I (Comma Errors)	10.47	7.98	7.04	32.76%
All Errors Combined[2]	15.58	12.13	9.81	36.05%
Standard Deviation				
E	1.65	1.18	.97	
F	1.29	.92	1.21	
G	.87	.48	.47	
H	3.71	2.84	2.12	
I	6.17	5.07	4.24	
All Errors Combined	9.37	7.55	5.53	
Maximum				
E	8	5	4	
F	5	4	5	
G	3	2	2	
H	15	17	9	
I	31	19	16	
All Errors Combined	51	36	24	

In addition to the average number of errors declining across the categories, the second panel shows that the variations, measured by standard deviation, among students have also declined. Coupled with the reduction in average, the decline in variation implies that students who used to make many errors were no longer doing so by the end of the semester. This trend is also shown in the third panel, which reports the maximum number of errors in each category.

From Assignment 1 to 4, there was over 30% reduction in errors in all categories, with the smallest change in category I (Comma Errors) at 32.76% and the largest change in category G (Non-capitalization) at 70.59%.

It is instructive to compare our results with the findings in Lunsford and Lunsford (2008), hereafter LL2008, wherein the authors identify the 20 most common errors in a nation-wide sample of 877 student papers.[3] Our writing assignments are similar to LL2008 in length. LL2008 estimated an average length of 1,038 words, while the average in our sample is around 900 words. Based on Table 7 in LL2008 (p. 795), we mapped the 20 categories of errors considered in their study into the five categories in ours and calculated the percentage and average number of errors. The calculations are shown in columns (4) and (5) of Table 2. We also calculated the percentages in the first three assignments in our studies, and those are reported in columns (1) through (3) in Table 2. Lastly, column (6) in Table 2 is the same as the means for Assignment 4 in Table 1, which is comparable to the numbers in column (5) from LL2008.

Several observations are worth mentioning. First, our study is similar to LL2008 in category E, Sentence Fragments. About 4% of the errors are in this category, and on average, about 0.5 mistakes were made in each assignment. The average of category E in A4, 0.43, is not statistically different from the 0.61 calculated from LL2008 in column (5) at 5% significance level. Based on the standard deviation reported in Table 1 and a sample size of 47, which is the number of students who turned in Assignment 4, we can test the null hypothesis of the equality of the two values 0.43 and 0.61. The p-value for the alternative

[2] "All Errors Combined" in Table 1 refers to the total number of errors students made in each submission. In the calculation of the means (i.e., 15.58, 12.13, 9.81) the average total errors per student is equal to the summand of the averages of each category because the total number of errors an average student makes is mathematically equivalent to the summand of errors this average student makes in each category. This is true based on the distributive law of multiplication. The distributive law does not hold for standard deviation and maximum because they are not linear combinations of different categories in their computation.

[3] Appendix B shows the full mapping of the "Credibility Killers" categories to the top 20 LL2008 categories.

Table 2
Comparison with Lunsford and Lunsford (2008)

	(1) A1	(2) A3	(3) A4	(4) Lunsford & Lunsford (2008) %	(5) Average	(6) A4 Average
E	4.85%	4.03%	4.34%	4.29%	.61	.43
F	4.99%	4.58%	4.99%	10.55%	1.5	.49
G	3.28%	2.01%	1.52%	9.30%	1.33	.15
H	19.69%	23.44%	19.09%	36.14%	5.17	1.87
I	67.19%	65.75%	71.80%	39.71%	5.69	7.04
Total					14.3	9.98

hypothesis of the two-sided test is .204, which means that the null hypothesis of equality is not rejected at the traditional 5% level of significance.

Second, students in our sample made significantly fewer errors in categories G (Non-capitalization) and H (Misspelled Words) than those in LL2008. We postulate that this has to do with the content of the writing. In our assignment, most capitalization would be company names, and since students' submissions were based on sample cases, they are unlikely to miss too many of these capitalizations. As for misspelled words, improvements in spell check functions in software such as Microsoft Word may be the reason behind improvement from LL2008, which was done in 2006, to our study.

Third, while we were concerned that our students made significantly more errors in category I, Comma Errors, a comparison between columns (5) and (6) reveals that the average numbers per assignment (or per about 1,000 words) are not significantly different from those in LL2008. During our assessment meetings, a question arose as to how Run-on Sentences and Comma Errors are being distinguished during grading. It seemed like it was unclear as to how our grader has distinguished between the two. Interestingly, since our students performed similarly to LL2008 in category E (Sentence Fragments), there is good reason to believe that our students may also perform similarly in category F (Run-on Sentences). The gap in category F between the two studies is about 1 error per assignment. If that is the error margin between categories E and I, then the number of comma errors in our assignments should be around 6 (1 fewer from 7.04) which is very similar to 5.69 as reported in LL2008.

When assessing writing, it may be misleading to think in terms of "how many" errors or types of errors are made. Instead, an effective writing sample should have no or only a very small number of errors.

When assessing writing, it may be misleading to think in terms of "how many" errors or types of errors are made. Instead, an effective writing sample should have no or only a very small number of errors. Table 3 identifies the percentage of students with one or fewer errors in a given category in the students' submissions for Assignments 1, 3, and 4. The four panels of Table 3 correspond to the following cases: (1) Full sample; (2) Full sample but treated a single error as no error; (3) Sub-sample with the 30 students who submitted all writing assignments; (4) Sub-sample and treated a single error as no error.

The improvement in category F (Run-on Sentences) depends on which panel of Table 3 we consider. The percentage of students without run-on sentence errors improves 9.36 percentage points for the entire set when a single error is marked as no error (Panel 2) and 26.67 percentage points when analyzing the 30 students who completed all four assignments (Panel 3).

For category G (Non-capitalization), the percentage of students without capitalization errors improves 11.3 percentage points for the entire set when a single error is marked as no error (Panel 2) and 20.47 percentage points in the full sample. Moreover, among the students who turned in all four writing assignments, by the last assignment, none of the students had made more than one mistake in category G (Non-capitalization).

Individual Analysis (n=30)

While aggregate results may have shown improvement throughout a semester, it may be more informative to look at individual data for the 30 students who completed

Table 3
Percentage of Students with One or Zero Errors, by Category and Assignment

	A1	A3	A4
Full sample with 0 error			
E (Sentence Fragments)	71.11%	80.00%	76.60%
F (Run-On Sentences)	62.22%	64.44%	78.72%
G (Capitalization)	68.89%	77.78%	89.36%
H (Misspelled Words)	31.11%	17.78%	31.91%
I (Comma Errors)	0	4.44%	2.13%
Full sample with 1 or 0 error			
E	82.22%	86.67%	91.49%
F	80.00%	86.67%	89.36%
G	84.44%	97.78%	95.74%
H	35.56%	31.11%	51.06%
I	0	6.67%	4.26%
30 student sample with 0 error			
E	70.00%	80.00%	73.33%
F	53.33%	63.33%	80.00%
G	73.33%	80.00%	93.33%
H	23.33%	10.00%	30.00%
I	0	3.33%	3.33%
30 student sample with 1 or 0 error			
E	80.00%	86.67%	90.00%
F	73.33%	83.33%	86.67%
G	86.67%	96.67%	100.00%
H	26.67%	20.00%	40.00%
I	0	6.67%	3.33%

all four assignments. In Figure 1, we plot the total number of mistakes in Assignment 1 (X-axis) and Assignment 4 (Y-axis) and examine their correlation. A fitted line (dash) that is less than 45-degree sloped shows that overall the total number of mistakes has declined. The fitted line plots the predicted number of Assignment 4 total errors resulted from a linear regression of Assignment 4 total errors (dependent variable) on Assignment 1 total errors (independent variable). We also show the 45-degree line (solid) in Figure 1. A dot on the 45-degree line indicates that a student has made the same number of mistakes in the two assignments, while a dot above/below indicates more/fewer mistakes. Only six students made more mistakes in the last assignment compared to the first.

Figure 1
Comparing Assignments 1 and 4 (n=30)

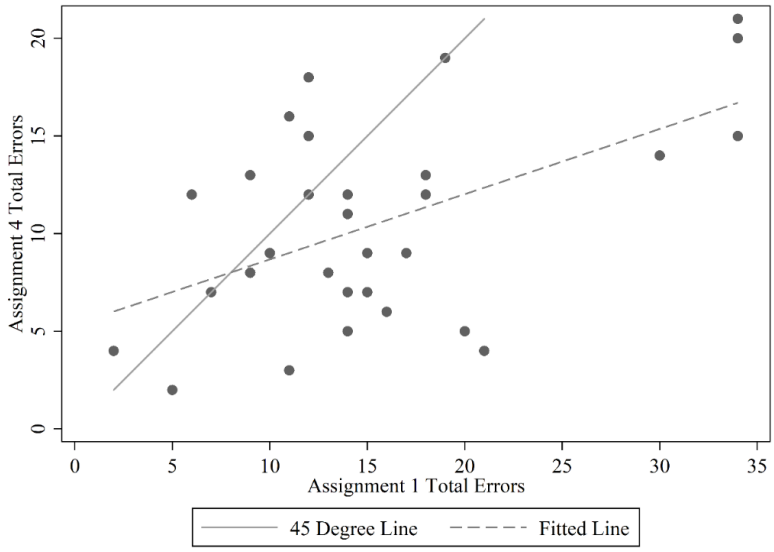
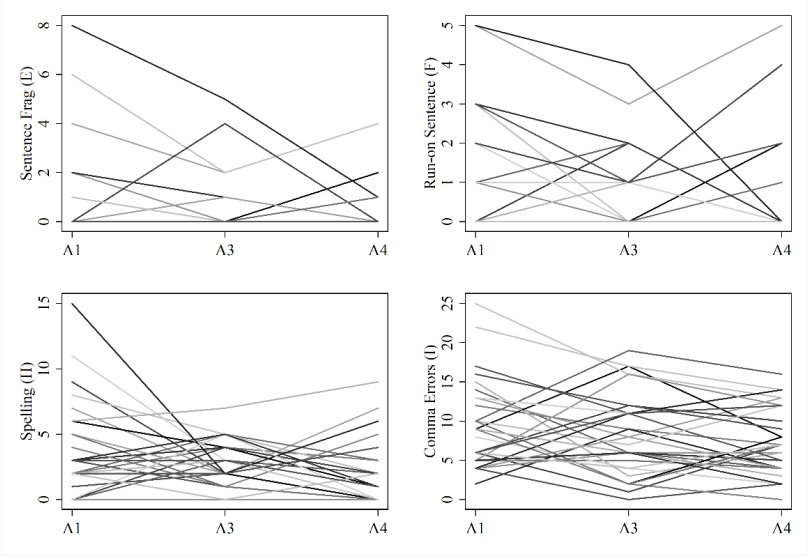
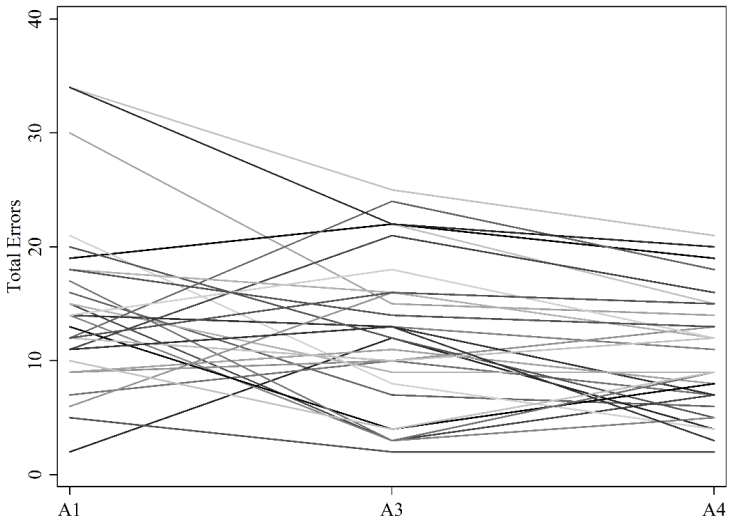


Figure 2
Movements in Errors in Categories E, F, H, and I (n=30)



Only Assignments 1, 3, and 4 are shown. Capitalization (G) is skipped.

Figure 3
Movements in Total Number of Errors (n=30)



In Figure 2, we plot the movements in errors in categories E, F, H, and I (Sentence Fragments, Run-on Sentences, Misspelled Words, Comma Errors) for the 30 students who completed all four assignments. Category G (Non-capitalization) is skipped in Figure 2 because its number tended to be small and invariable. As mentioned before, Assignment 2 is skipped due to peculiar behaviors.

Figure 2 suggests that students could have taken the previous WI feedback as a signal rather than a tool of learning and improvement. Many students' error counts exhibit a *V* or *inverted-V* shape. In other words, students who made many mistakes in the first assignment tended to make fewer mistakes in the third assignment but rebounded in the fourth assignment—indicated by a V-shaped line. In contrast, students who made few mistakes in the first assignment tended to make more mistakes in the third assignment but improved in the fourth assignment. A similar phenomenon can be seen also in the total number of errors as shown in Figure 3.

Table 4 summarizes the number of students who exhibited V or inverted-V movements in categories E, F, H, I, as well as in all errors combined. Note that the All Errors row is not an aggregation of the four categories.

Table 4
Summary of Inverted- and V-shape Patterns

	V Shape	Inverted-V Shape	Total
E (Sentence Fragments)	5	2	7
F (Run-On Sentences)	6	5	11
H (Misspelled Word)	6	13	19
I (Comma Errors)	9	7	16
All Errors Combined	5	4	9

Table 4 shows that the frequency of V shape versus inverted-V shape responses is similar for category F and I, as well as for total errors. However, there are many more V-shaped patterns in category E and many fewer V-shaped patterns in category H.

The assessment of the WI's grading results shows that, on average, a group of students can decrease their frequency of errors over a semester (as shown in Table 1).

While it is encouraging that students generally reduced their errors across the assignment sequence, several questions are raised by students' V and inverted-V shaped trend patterns. There are many potential explanations for such trends, including but not limited to cognitive overload (Bean, 2011), individual proclivities (O'Neill, 2018), or exacerbation of writing anxiety (Ekholm et al., 2015). The following section further discusses potential implications and connections to existing research on students' writing skills and skill development.

Discussion

The WI encourages faculty to incorporate writing assignments into their courses by providing some assessment support to the faculty regarding students' grammar and punctuation usage. The results provide insight into the potential influence of the WI on the participating students' performance as a group, providing implications about the overall effectiveness of the program on reducing the occurrence of errors in students' writing. The results also provide some insight into the grammar errors that the students make, tend to avoid, and have the most opportunity to improve.

Broad Trends in Error-Making

The assessment of the WI's grading results shows that, on average, a group of students can decrease their frequency of errors over a semester (as shown in Table 1). Also, generally, the proportion of students including each error type decreases between Assignments 1 and 4, showing that most students are reducing their common grammar and punctuation errors over time. The overall decrease in error frequency between the first and last assignments suggests that the WI may support the improvement of students' grammar and punctuation usage. Although the feedback is limited, the feedback from the WI grader does provide students with some grammar instruction that is contextualized within their writing, following research-based recommendations (e.g. Lancaster & Olinger, 2014; Myhill et al. 2013). However, research showing the value of teaching grammar in context typically refers to instruction that involves more student-instructor discussion and student revision of their work. Studies testing the influence of corrective feedback that is accompanied by little to no explanatory feedback showed less long-term effectiveness in improving students' grammar usage (Bitchener & Knoch, 2010; Yu et al., 2020). In this study, the WI grader did not have direct contact with the students to discuss their feedback, an area of attention for future discussions.

The trends between the full sample of 54 students and the sub-sample of 30 students who completed all four assignments are similar in that both sample sets showed a general reduction in errors between Assignments 1 and 4. However, an examination of the sub-sample offers more nuanced implications about the influence of the WI feedback. Of the 30 students who submitted all four assignments, 21 decreased their total number of errors between the first and last assignments (70%). When tracking the frequency of errors across assignments, some of the sub-sample students increased their total number of errors between either Assignments 1 and 3 or Assignments 3 and 4. These spikes, visualized previously by the V and inverted-V lines in Figures 2 and 3, raise questions about the influence of the WI feedback.

One assumption of the WI is that students will see the frequency and type of errors they include in their writing, learn the correct way to avoid such errors, and then reduce or eliminate the error(s) in the future. It is possible that students with V-shaped trend lines took a signal from the WI feedback and reduced their errors in the third assignment, but then, when completing the fourth and final assignment, faced some other challenge. Students may have experienced cognitive overload due to the challenge to both avoid grammatical errors and master Assignment 4's required content knowledge. Schwalm's landmark study revealed that "grammatical competence begins to drop off as the tasks become more complex" (cited in Bean, 2011, p. 77). Additionally, extensive or unclear corrective feedback can also trigger students' writing anxiety, which can increase errors that students might otherwise be able to avoid (Bean, 2011; Ekholm et al., 2015; Mascle, 2013; Zumbrunn et al., 2016). Ultimately, without qualitative feedback from the students, it is unclear how the students used the WI feedback in this situation, and further investigation would be useful.

The results show that the four Status Marking Errors were nearly absent from the entire sample: all students avoided using nonstandard verb forms and double negatives across all four assignments.

Rule-Specific Error Trends

The results show that the four Status Marking Errors were nearly absent from the entire sample: all students avoided using nonstandard verb forms and double negatives across all four assignments. There was only one "Subject-Verb Agreement" error in Assignment 2 and one "Object Pronoun as Subject Usage" error in Assignment 3. While research shows that these types of grammar errors have a negative influence on a reader's impression of the writer (Boettger & Emory Moore, 2018; Gray & Heuser, 2003; Gubala et al., 2020), these students seem to know to avoid these errors, at least when writing. The absence of these errors is mirrored in studies testing the 20 most frequent errors in college student writing. While Connors and Lunsford (1988) found "wrong verb form" ranked 13th and subject-verb disagreement ranked 14th, none of these four Status Marking Errors appeared in the top 20 most common errors in Lunsford and Lunsford's (2008) study.

Most students also showed an improved ability to avoid three other grammar errors from Assignment 1 to Assignment 4: sentence fragments, run-on sentences, and capitalization errors. The percentage of student papers that included one or fewer of these errors increased between Assignments 1 and 4, where:

- Avoiding sentence fragments increased 9.27 percentage points from 82.22% to 91.49%,
- Avoiding run-on sentences increased 9.36 percentage points from 80.00% to 89.36%
- Appropriate capitalization increased 11.30 percentage points from 84.44% to 95.74%

These improvement trends show that while 80% or more of students were successfully avoiding these errors at the beginning of the semester, students were still able to improve their ability to avoid these errors.

Based on these trends, faculty might expect most students to understand the rules regarding these grammar constructions, suggesting that additional support and instruction on these aspects are not necessary. However, we urge caution, as these results show only that these students improved their avoidance of the errors, rather than their understanding of the rule. Studies of professionals' perceptions of these errors show that sentence fragments and run-on sentences are among the most bothersome and serious errors (Boettger & Emory Moore, 2018; Gray & Heuser, 2003; Gubala et al., 2020; Hairston, 1981). Moreover, Lunsford and Lunsford (2008) found that sentence fragments were ranked as the 20th most common error and that the 15th and 16th most common errors were both versions of run-on sentences (fused sentences and comma splices, respectively). Thus, even while students in this study generally avoided these errors, this research on perceptions of error and error frequency shows the continued importance of discussing these rules with students.

Interestingly, students showed the least improvement in avoiding misspellings and comma errors. Students demonstrated a stronger ability to avoid spelling-related errors than comma errors, reducing the number of misspelled words between Assignments 1 and 4. In

the full sample, the percentage of students who included one or fewer misspelled words increased by 15.5 percentage points from 35.56% to 51.06%. This trend is promising, yet only about half of the students were able to include one or fewer spelling errors in their assignments. Error-marking by the WI grader conflated a variety of spelling errors into this category, including general misspellings, "typos," homophone errors, and wrong word errors. Research on professionals' perceptions of errors typically separate spelling-related errors into different categories, but multiple spelling-related errors have been identified as bothersome or serious errors. "Wrong word" errors are among the most bothersome, and general misspelling errors typically appear near the middle of "most bothersome" lists (Boettger & Emory Moore, 2018; Gray & Heuser, 2003; Gubala et al., 2020).

Interestingly, students showed the least improvement in avoiding misspellings and comma errors. Students demonstrated a stronger ability to avoid spelling-related errors than comma errors, reducing the number of misspelled words between Assignments 1 and 4.

Despite the negative perceptions of spelling-related errors, Lunsford and Lunsford (2008) found that "wrong word" errors and spelling errors, including homonym errors, are the first and fifth most common errors that appear in their study of student writing. In Connors and Lunsford's (1988) study of common student errors, the only spelling-related error in the top twenty errors was "wrong word" errors, ranked as fourth most frequent, suggesting a potential increase in spelling-related errors. Our results and the corresponding research raise concerns about students' inclusion of spelling-related errors. Lunsford and Lunsford (2008) found that many of the "wrong word" errors in their study seemed to be due to auto-corrections or spell-checker suggestions offered by word processing programs. Nevertheless, in Gubala et al.'s (2020) study, even though misspellings received a mean "bothersome" score of 2.94/5.0 from professionals, the comments regarding misspellings indicated a strong negative evaluation of the writer's intelligence, care, and competence. Thus, the improvement shown by students in this study is promising, but students' spelling-related errors continue to be an area for improvement.

The most recurring errors and least change in error frequency occurred in the Comma Errors category (category I). Nearly every submission in all four assignments included comma errors. Only 4.26% of the full sample included one or fewer comma errors by Assignment 4. Still, this small percentage is an improvement since 0% of the submissions in Assignment 1 included one or fewer comma errors. These results may not be particularly surprising given that in Lunsford and Lunsford's (2008) study, four of the top twenty errors are comma usage errors.

As a whole, students showed notable improvement in avoiding spelling-related errors, even if only half of the final assignments included one or fewer of these errors. There were practically no Status Marking Errors in the full set of writing samples, and most students showed an ability to generally avoid using sentence fragments, run-on sentences, and inappropriate capitalization. Though comma usage seems to be an area of deficiency for these students, we discuss in the following section potential limitations of our study based on WI grading protocols.

Implications and Conclusion

Overall, the preliminary analysis of grading reports from the Writing Initiative showed some positive trends for upper-level business students. However, limitations and questions emerged during the analysis that suggest the need for additional research. As previously mentioned, the comma rules category of the rubric includes only two usage rules, though we know the grader identified additional comma errors in students' writing. Also, comma-splice errors may have been marked inconsistently in different categories, thereby skewing the results. The results might also have been influenced by variance in grammar instruction. The business communication standard is to use the Oxford or Serial Comma (the last comma in a series; e.g., I enrolled in Economics, Marketing, and Communication courses). Some students, however, may have been taught to not use this comma because in some industries, it is considered optional. Moreover, professional and academic evaluators have shown significant inconsistencies in their marking of grammar errors (e.g., Gray & Heuser, 2003; Lunsford & Lunsford, 2008). Therefore, this variance might cause difficulties for students learning "correct" comma usage and in our assessment of the frequency and correction of those errors consistently across the disciplines.

Additionally, we acknowledge that grammar and punctuation are only two aspects of effective written communication. Grading for grammar and punctuation alone could send the signal—to students and other faculty—that the overall argument, organization, cohesiveness, and clarity of the writing sample are less important. The feedback procedure used in the Writing Initiative, wherein the WI grader indicates every error and type and rarely includes developmental explanations, could also be overwhelming for students. This situation is likely compounded by the lack of guidance for faculty about how to discuss writing skills and the WI with students, since the administration's messaging to faculty focuses on grading percentages and implementation logistics. Though our results indicate a potential need for increased classroom intervention towards the end of the semester, we do not know definitively what contributed to the resurgence in error-making. Further study on students' perceptions of the feedback would better illuminate how they are using the marks from the grader to improve their writing. These efforts would also allow for more precise pedagogical interventions and assessment measures.

In our college, we plan to continue addressing these questions through our assessment efforts. The college's communication assessment committee, on which the authors serve, assesses the majors' communication skills once every two years. This article presents the results of the college's first assessment of the WI in its current format. Based on these preliminary results, and other communication assessment results, we plan to implement three adjustments that may influence the WI. First, the curriculum in the college's business communications courses will be adjusted to reduce focus on the errors that students do not make as often and increase developmental attention on areas in which students show deficiencies, including grammar and other areas. Second, the college administration and business communication faculty are collaborating to develop a better rubric for the WI that will provide further clarity to both students and the graders. The updated rubric will parse categories further like those in Lunsford and Lunsford (2008). The college's communication assessment committee also plans to investigate how students use the WI grader feedback, which will provide more robust insight into the initiative's effect. Last, a plan is being developed to track and assess students' progress in writing through their academic careers. This plan will involve collecting data for multiple years in both lower- and upper-level classes that have a writing component. Through such efforts from the college administration, assessment committees, and faculty, our college aims to refine students' communication skills and align programmatic efforts with best practices.

…We acknowledge that grammar and punctuation are only two aspects of effective written communication. Grading for grammar and punctuation alone could send the signal—to students and other faculty—that the overall argument, organization, cohesiveness, and clarity of the writing sample are less important.

Based on these preliminary findings, we encourage organizers of support programs like this one to consider conducting preliminary research to customize grading criteria focused on the most common writing concerns for their student population, which might account for more than just grammar and mechanics. For example, the inclusion of the Status Marking Errors from the Credibility Killers rubric might have been unnecessary for our students, based on the 174 assignments included in this sample. The Writing Initiative grader marked only 2 errors in the Status Marking Errors category (nonstandard verb forms; lack of subject-verb agreement; use of double negatives; and object pronoun as subject) across all submissions. While these errors are important ones to avoid, the rarity with which these errors appear in student writing suggests that students', graders', and faculty members' time would be better spent on the more frequent error types. Removal of these infrequent errors may provide an opportunity to include other more frequent types of errors, potentially leading to clearer and more professional writing from our students. Moreover, Boettger and Emory Moore's 2018 study of professionals' evaluation of errors found that half of the errors that participating professionals indicated were most bothersome related to design rather than word usage. We suggest a more focused approach to feedback that is based on timely research and institutional evidence of students' most frequent errors, plus support through classroom instruction that addresses common errors identified by the grader. Nevertheless, using a support program like the Writing Initiative may incentivize students to take more care with the "correctness" of their writing, potentially leading to increased competency in written communication.

Finally, when considering a similar writing support initiative, organizers should concretize the purpose of the program. Our Writing Initiative's stated purpose is to develop student writing but, in fact, it seems to function in a way that is focused on faculty and encouraging them to integrate writing into their discipline-based courses. If the program was to be truly student-focused, we would consider modifications in both the application and assessment of the grading feedback, to ensure alignment between the purpose of a writing support program and the practices integrated to achieve that purpose.

References

Addams, L. & Allred, A. (2015). Business communication course redesigned: All written and oral communication assignments based on building career skills. *Academy of Educational Leadership Journal, 19*, 250-265. https://search.proquest.com/openview/c2af477dff4fa4a8f8d150b9b764918a/1?cbl=38741&pq-origsite=gscholar

Austin, T., Sigmar, L. Mehta, G. & Shirk, J. (2018). The impact of web-assisted instruction on student writing outcomes in business communication. *Journal of Instructional Pedagogies, 20*, 1-11. https://files.eric.ed.gov/fulltext/EJ1178728.pdf

Bean, J. C. (2011). *Engaging ideas: The professor's guide to integrating writing, critical thinking, and active learning in the classroom* (2nd ed.). Jossey-Bass.

Beason, L. (2001). Ethos and errors: How business people react to errors. *College Composition and Communication, 53*, 33-64. https://www.jstor.org/stable/359061

Bitchenor, J. & Knoch, U. (2010). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing, 19*(4), 207-217. https://doi.org/10.1016/j.jslw.2010.10.002

Boettger, R. K. & Emory Moore, L. E. (2018). Analyzing error perception and recognition among professional communication practitioners and academics. *Business and Professional Communication Quarterly, 81*(4), 462-484. https://doi.org/10.1177%2F2329490618803740

Caldwell, C. & Al-Ajmi, N. J. H. (2018). Improving business student writing – A ten-step model. *Business and Management Research, 7*(1), 35-41. https://doi.org/10.5430/bmr.v7n1p35

Carpenter, R., Whiddon, S., & Dvorak, K. (2014). Guest editor introduction: Revisiting and revising course embedded tutoring facilitated by writing centers. *Praxis: A Writing Center Journal, 12*, 3-7. http://www.praxisuwc.com/from-the-guest-editors

Connors, R. J. & Lunsford, A. A. (1988). Frequency of formal errors in current college writing, or Ma and Pa Kettle do research. *College Composition and Communication, 39*(4), 395-409. https://www.jstor.org/stable/357695

Ekholm, E., Zumbrunn, S., & Conklin, S. (2015). The relation of college student self-efficacy toward writing and writing self-regulation aptitude: Writing feedback perceptions as a mediating variable. *Teaching in Higher Education, 20*(2), 197-207. https://doi.org/10.1080/13562517.2014.974026

Gray, L. S. & Heuser, P. (2003). Nonacademic professionals' perception of usage errors. *Journal of Basic Writing, 22*(1), 50-70. https://wac.colostate.edu/jbw/v22n1/gray.pdf

Grouling, J. (2018). The genre of teacher comments from hard copy to iPad. *Journal of Response to Writing, 4*(1), 70-99. https://journalrw.org/index.php/jrw/article/view/103

Gubala, C., Larson, K., & Melonçon, L. (2020). Do writing errors bother professionals? An analysis of the most bothersome errors and how the writer's ethos is affected. *Journal of Business and Technical Communication, 34*(3), 250-286. https://doi.org/10.1177/1050651920910205

Hairston, M. (1981). Not all errors are created equal: Nonacademic readers in the professions respond to lapses in usage. *College English, 43*(8), 794-806. http://www.jstor.org/stable/376679

Keane, K., McCrea, D. & Russell, M. (2018). Personalizing feedback using voice comments. *Open Praxis, 10*(4), 309-324. International Council for Open and Distance Education. https://www.learntechlib.org/p/207182/

Lancaster, Z. & Olinger, A. R. (2014). *Teaching grammar-in-context in college writing instruction: An update on the research literature, WPA-CompPile Research Bibliographies, No. 24*. WPA-CompPile Research Bibliographies. http://comppile.org/wpa/bibliographies/Bib24/Grammar_in_Context.pdf

Lunsford, A. A. & Lunsford, K. J. (2008). "Mistakes are a fact of life": A national comparative study. *College Composition and Communication, 59*(4), 781-806. https://www.jstor.org/stable/20457033

Masele, D. D. (2013). Writing self-efficacy and written communication skills. *Business Communication Quarterly, 76*(2), 216-225. https://doi.org/10.1177%2F1080569913480234

Myhill, D. A., Jones, S. M., Lines, H., & Watson, A. (2013). Re-thinking grammar: The impact of embedded grammar teaching on students' writing and students' metalinguistic understanding. *Research Papers in Education, 27*(2), 139-166. https://doi.org/10.1080/02671522.2011.637640

National Association of Colleges and Employers (NACE). (2018). Job Outlook 2019 Survey. http://www.naceweb.org

O'Neill, K. S. (2018). Applying the Pareto Principle to the analysis of students' errors in grammar, mechanics, and style, *Research in Higher Education Journal, 34*, 1-12. https://eric.ed.gov/?id=EJ1178476

O'Neill, K. & Sigmar, L. (2014, October 22-25). *Using communication processes to accelerate AACSB improvement progress* [Conference presentation]. Association for Business Communication 79th Annual International Convention, Philadelphia, PA, United States.

Quible, Z. K. (2006). Impact of error labeling on error elimination in business writing. *Business Communication Quarterly, 69*, 8-24. https://doi.org/10.1177/1080569905285597

Sigmar, L. S. & Hynes, G. E. (2012). Major difference: An examination of student writing performance by major and its implications for business communication. *Administrative Issues Journal: Education, Practice, and Research, 2*(1), 133-45.

Sigmar, L. & Austin, T. (2013, October 23-26). *Impression management: The professionals' guide to relevant grammar* [Conference presentation]. Association for Business Communication 78th Annual International Convention, New Orleans, LA, United States. https://eric.ed.gov/?id=EJ1056394

Sigmar, L. & Austin, T. (2015, March 11-14). *The professionals' guide to relevant grammar: The sequel* [Conference presentation]. Association for Business Communication-Southwest Region Annual Conference, Houston, TX, United States.

Willis, L., Wilkie, L., & Gracey, D. M. (2012). An analysis of the grammar skills of undergraduate business students. *Global Education Journal, 2012*(4), 56-80. https://www.franklinpublishing.net/globaleducation.html

Yu, S., Jiang, L., & Zhou, N. (2020). Investigating what feedback practices contribute to students' writing motivation and engagement in Chinese EFL context: A large scale study. *Assessing Writing, 44*. https://doi.org/10.1016/j.asw.2020.100451

Zumbrunn, S., Marrs, S., & Mewborn, C. (2016). Toward a better understanding of student perceptions of writing feedback: A mixed methods study. *Reading & Writing, 29*(2), 349-370. https://link.springer.com/article/10.1007/s11145-015-9599-3

Abstract

A grand challenge is a problem that requires broad cooperation for successful resolution from a community of scholars. Several national and international organizations have generated lists of grand challenges to unify the efforts of scholars and practitioners in a field. However, the field of assessment has yet to identify its own set of grand challenges that could serve to organize and motivate progress toward meaningful goals. This article describes the process by which potential grand challenges were identified and subsequently evaluated by professionals in the field through a national survey. Results of the survey demonstrate broad support for the importance of four challenges: 1) Use assessment findings to increase equity; 2) Use assessment findings to direct immediate pedagogical improvements; 3) Produce visible and actionable assessment findings that drive innovation; and 4) Examine changes in institutional effectiveness (including student learning) over time. The article concludes with a discussion of the grand challenges that emerged from this work and a description of an ongoing national effort to address these challenges through strategic planning.

AUTHORS

Karen E. Singer-Freeman, Ph.D.
University of North Carolina at Charlotte

Christine Robinson, Ph.D.
University of North Carolina at Charlotte

Grand Challenges for Assessment in Higher Education

A grand challenge is a problem that requires broad cooperation for successful resolution from a community of scholars. Hilbert (1902) first identified grand challenges by publishing a list of mathematics problems with the goal of advancing solution creation. Since that time, national and international organizations have generated lists of grand challenges to unify the efforts of scholars and practitioners at research universities, federal agencies, and non-profit organizations (Omenn, 2006; Popowitz & Dorgelo, 2018; Uehara et al., 2014; Varmus et al., 2003). The articulation of grand challenges has proven useful as a means of creating synergistic research efforts to make a positive difference in the world. Examples of effective grand challenges include creating economical sources of solar energy (National Academy of Engineering, 2016), developing renewable fuel alternatives (National Research Council, 2005), and including active science inquiry in all introductory college science classes (Alberts, 2013). However, the field of assessment has yet to identify its own set of grand challenges that could serve to organize and motivate progress. In this study, we sought to identify compelling grand challenges for the field of higher education assessment.

CORRESPONDENCE

Email
ksingerf@uncc.edu

Why Grand Challenges?

There is a pressing need to improve perceptions about the value of assessment in higher education. In a recent survey of chief academic officers, nearly a third believed their college's assessment efforts were more about keeping politicians and accreditors happy than improving teaching and learning, and nearly a fifth disagreed that assessment

systems have improved the quality of teaching and learning (Jaschik & Lederman, 2020). Assessment professionals report disliking the need to persuade others about the value of assessment (Ariovich et al., 2018). Recent surveys of assessment professionals have focused primarily on demographics, salary, and job responsibilities (Ariovich et al., 2018; Combs & Rose, 2016; Nichols & Slotnick, 2018). To date, there has not been a survey of assessment professionals' beliefs about important future directions for the field. The identification of grand challenges described in this paper served as the starting point for national strategic planning, in which the assessment field will coordinate research and practical efforts to increase the use of assessment findings for improvements in teaching and learning. This national planning effort, which is currently underway, will strengthen the commitment of higher education leaders by improving the quality of assessment and publicizing the positive impact of quality assessment. Strong leadership buy in is essential to increase use of assessment findings in data-driven decision-making (Banta et al., 2016).

This national planning effort, which is currently underway, will strengthen the commitment of higher education leaders by improving the quality of assessment and publicizing the positive impact of quality assessment.

This article describes the process by which potential grand challenges were identified through a national survey of higher education assessment professionals. The survey was conducted by the authors, without support from any organizations. However, since the completion of the survey, the Grand Challenges in Assessment Project has received endorsements from nine national organizations (American College Personnel Association, Indiana University-Purdue University Indianapolis Assessment Institute, Association for Institutional Research, Association for the Assessment of Learning in Higher Education, Association of American Colleges and Universities, Council for the Advancement of Standards in Higher Education, National Association of Student Personnel Administrators, National Institute of Learning Outcomes Assessment, and Student Affairs Assessment Leaders). The paper concludes with a discussion of the grand challenges that emerged from this project and a description of the Grand Challenges in Assessment Project.

Method

Different fields have approached the selection of grand challenges in different ways, including relying on the work of a single individual, reviewing current literature, holding symposia, or issuing broad calls for proposals (Gould, 2010). For the development of the survey, the authors reviewed current assessment literature and selected challenges that were mentioned frequently. The four characteristics used to identify grand challenges were: (1) Is extremely hard to do, yet doable; (2) Would produce positive outcomes potentially affecting large numbers of people; (3) Is associated with clear metrics and goals so progress and completion can be identified; and (4) Would capture popular imagination, and thus garner political support (Gould, 2010; Stephan et al., 2015). To identify the challenges for inclusion in the survey, the authors reviewed assessment websites, blogs, discussion boards, and publications from 2015 to 2019. Focusing on publications within that specific time frame, rather than conducting a more extensive review, maintained a future-oriented perspective for challenge identification. We chose to include peer-reviewed and non-peer-reviewed publications, blogs, and discussion boards in our review to identify challenges that were generating practical interest, as well as challenges that were discussed in the published literature. In total, we reviewed 83 pieces of writing that included 46 non-peer-reviewed sources, 34 peer-reviewed sources, and 3 blog or discussion board posts. The total number and distributions of materials reviewed is shown in Table 1.

As we read each source, we noted all challenges facing the field of assessment that fulfilled the four defining characteristics of the grand challenges described above. The review resulted in the identification of 10 potential challenges, described in Table 2. Most sources referenced more than one of the 10 challenges that were identified. As shown in Table 2, all challenges were referenced in multiple sources. A full description of these challenges can be found in Singer-Freeman and Robinson (2020).

Survey Instrument, Sampling, and Administration

Drafted survey items, which emerged from the literature review, were shared with one assessment professional employed at a community college, one college administrator at a four-year college, and five staff members from a national assessment organization. These

Table 1
Sources Reviewed to Identify Grand Challenges in Assessment Listed in Order of Frequency

Source	References
34 peer-reviewed sources	
Research & Practice in Assessment	6
Assessment & Evaluation in Higher Education	4
Intersection	4
Emerging Dialogues	3
Educational Assessment, Evaluation and Accountability	2
International Journal of ePortfolio	2
Planning for Higher Education	2
Assessment in Education: Principles, Policy, & Practice	1
British Journal of Educational Psychology	1
Journal of Competency-Based Education	1
Educational Planning	1
Journal of Biochemistry and Molecular Biology Education	1
Journal of General Education	1
Journal of Higher Education	1
New Directions for Evaluation	1
Online Learning Journal	1
Journal of Teaching and Learning	1
Urban Education	1
46 non-peer-reviewed sources	
Assessment Update	13
Book Chapters	8
Liberal Education	6
NILOA publications and Occasional Papers	6
AAC&U Reports	5
Viewpoints	3
American Council on Education	1
Inside Higher Education	1
London: Higher Education Commission	1
Lumina Issue Paper	1
Pell Institute for the Study of Opportunity in Higher Education	1
3 blogs and discussion board posts	
Linda Suskie Blog	2
Educause Review	1

individuals were selected to represent a range of institutional perspectives on assessment. The items were revised in response to their feedback. The final survey contained 44 4-point Likert scale questions (responses: "not at all" = 1, "a little" = 2, "to some extent" = 3, "very much" = 4). Respondents evaluated the extent to which each of the 10 identified grand challenges could be described using each of the four characteristics (Is extremely hard to do, yet doable; Would produce positive outcomes potentially affecting large numbers of people; Is associated with clear metrics and goals so that progress and completion can be identified; and Would capture the popular imagination, and thus garner political support). After evaluating the different challenges, respondents were asked to rank the overall importance of the 10 grand challenges. The survey also included two open-ended questions. The first asked for additions, deletions, or changes to the characteristics of grand challenges. The second asked participants to suggest an additional challenge, if so desired. The survey also included eight demographic questions. The complete survey is included in the Appendix.

For the development of the survey, the authors reviewed current assessment literature and selected challenges that were mentioned frequently… In total, we reviewed 83 pieces of writing that included 46 non-peer-reviewed sources, 34 peer-reviewed sources, and 3 blog or discussion board posts.

Invitations to complete the survey were shared on assessment listservs (ASSESS, AAHLE, SAA-Leaders), the authors' LinkedIn accounts, and in the National Institute for Learning Outcomes Assessment (NILOA) newsletter. The authors also sent email invitations to personal contacts and requested that all recipients distribute the survey invitation to others in the field. After providing informed consent, participants completed the survey. Initially, the survey was set to require responses to all questions. However, in

Table 2
Challenges Identified from Review of Recent Scholarly Work

Short title	Full challenge	Referenced
RELATED TO CONTINUOUS IMPROVEMENT		
1. DRIVE INNOVATION	Produce visible and actionable assessment findings that drive innovation.	22
2. INFORM BUDGET	Use assessment findings to inform budgetary initiatives	6
3. IMMEDIATE IMPROVEMENTS	Use assessment findings to direct immediate pedagogical improvements.	8
RELATED TO ADDRESSING INEQUITIES		
4. INCREASE EQUITY	Use assessment findings to increase equity.	12
5. DISAGGREGATE DATA	Disaggregate data to include important student characteristics.	8
RELATED TO IMPROVING MEASUREMENT		
6. CHANGE OVER TIME	Examine changes in institutional effectiveness (including student learning) over time.	11
7. STUDENT SELF-EVALUATION	Involve students in authentic self-evaluation of their learning.	8
8. ePORTFOLIOS	Use ePortfolios to capture students' learning over the entire span of their education.	11
9. MASSIVE DATA	Leverage technology to analyze massive data sets within and across institutions.	12
10. COMMUNICATE	Communicate relevant, timely, and contextualized information about student learning to stakeholders.	13

response to constructive feedback from several respondents in the first week of the launch, the questionnaire was reset to allow participants to skip any item. Participants spent between four and 76 minutes on the survey, with an average completion time of 14.66 minutes (*SD* = 10.73).

Participants

A total of 231 individuals submitted completed or partially completed surveys. An additional 176 individuals followed the link to the survey but did not complete or submit the survey. Because the survey was advertised broadly, it is difficult to calculate an accurate response rate. Of the individuals who followed the link, the response rate was 57%. A popular listserv that we used for distribution, ASSESS, is reported to have over 1,500 subscribers (Fuller et al., 2015). Assuming most assessment professionals we reached through other forms of outreach also subscribe to this listserv, that would result in a response rate of 15%. All submitted surveys were included in analyses, including those that were only partially completed. We received responses to the eight individual demographic questions from between 204 and 216 participants. Participants reported spending between two and 47 years employed in higher education with an average of 18.41 years (*SD* = 9.34). Participants reported spending between zero and 40 years involved in assessment activities with an average

of 11.73 years (*SD* = 8.28). Most participants (72%) reported working at public institutions of higher education, with less representation from other types of institution (27% private non-profit institutions and less than 2% for-profit institutions). Participants who selected multiple positions were included in all groups they selected. Most respondents reported working primarily in administrative roles (78%), with less representation from respondents in other roles (13% teaching, 12% research, 6% combined position, 3% professional organization, and 1% accrediting organization). Participants reported their gender as female (70%), male (29%), and genderqueer (1%). Participants who selected multiple race and ethnicity categories were included in all groups they selected. Participants reported their ethnicity and race as White (86%), Black or African American (10%), Asian (3%), Hispanic or Latino (2%), American Indian (2%), and Pacific Islander (less than 1%). Our sample was smaller than some other recent surveys (See Table 3); however, we believe our sample represents the assessment community and note that our sample shares similar demographics with other samples (Combs & Rose, 2016; Nichols & Slotnick, 2018).

Table 3
Comparison of Current Sample to Other Surveys of Assessment Professionals

	n	% primarily administrator	% primarily faculty or research	% White
Our Sample	231	78	23	86
Nichols & Slotnick (2018)	324	84	16	89
Ariovich et al. (2016)	1074	57	43	
Combs & Rose (2016)	377	81	5	
Frew et al. (2007)	203	> 66		

The challenge "Increase Equity" received the highest average score (M = 3.20), which was significantly higher than the average scores received for all other challenges, except "Drive Innovation" (M = 3.08) and "Change Over Time" (M = 3.05).

Results

Comprehensiveness of Characteristics Used to Assess Grand Challenges

When given the opportunity to suggest additions, deletions, or changes to the characteristics used to assess grand challenges (see above for characteristics), 177 participants (77%) provided no response or indicated that no changes were needed, and 54 (23%) provided suggestions. There were 34 suggestions of additional characteristics. One predominant theme emerged: Support sustainability of high-quality assessment by engaging broad participation or overcoming negativity (15 responses). In addition, there were 14 suggestions about ways to rephrase specific characteristics and seven concerns about whether the characteristics were a good fit for the field of assessment. Because data collection was complete, no actions were taken to alter the characteristics used to evaluate the challenges in response to these suggestions.

Extent to Which Grand Challenges Fulfill Each Characteristic

To investigate participant responses to the 10 challenges, we assigned each Likert response an ordinal score ("not at all" = 1, "a little" = 2, "to some extent" = 3, "very much" = 4) and calculated the average rating of the extent to which participants reported the four characteristics could be applied to each challenge. These scores are reported along with standard deviations in Table 4. Scores ranged from 1 to 4 on all items. An average score across the four characteristics of grand challenges measures the extent to which a challenge holistically exemplifies the characteristics of grand challenges. The use of parametric analyses has been deemed appropriate for aggregated Likert ratings (Harpe, 2015). To test for differences in the extent to which the 10 challenges met the characteristics of grand challenges, a single-factor within-subjects Analysis of Variance (ANOVA) was performed on the average rating for each challenge. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(44) = 141.96$, $p < .01$); therefore, degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = .90$). A significant difference was found between the 10 challenges, Wilks'$\chi^2 = .62$, $F(8.07, 1662.91) = 14.45$, $p < .001$ with a small effect size (Partial Eta Squared = .07). The results of post hoc comparisons using the Bonferroni correction are shown in Table 4. The challenge "Increase Equity" received the highest average score ($M = 3.20$), which was significantly higher than the average scores

received for all other challenges, except "Drive Innovation" (M = 3.08) and "Change Over Time" (M = 3.05). The challenge "ePortfolio" received the lowest average score (M = 2.73), which was significantly lower than the average scores received for all other challenges, except "Inform Budget" (M = 2.90). The remaining five challenges did not significantly differ from each other, with means ranging from 2.93 to 3.04.

Participants indicated the challenges most strongly fulfilled the characteristic "Would produce positive outcomes potentially affecting large numbers of people" (M = 3.23).

To test for differences in the extent to which the four characteristics of grand challenges were applied to the challenges, a single-factor within-subjects ANOVA was performed on the average rating for each characteristic. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(5) = 55.60$, $p < .01$), therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = .88$). A significant difference was found between the four challenges, Wilks' $\lambda = .53$ $F(2.64, 603.32) = 46.17$, $p < .001$ with a large effect size (Partial Eta Squared = .47). The results of post hoc comparisons using the Bonferroni correction are shown in Table 4. Participants indicated the challenges most strongly fulfilled the characteristic "Would produce positive outcomes potentially affecting large numbers of people" (M = 3.23). This characteristic received significantly higher ratings than the other three characteristics. Participants indicated the challenges least strongly fulfilled the characteristic "Would capture the popular imagination, and thus garner political support" (M = 2.77). This characteristic received significantly lower ratings than the other three characteristics. The characteristics "Is extremely hard to do, yet doable" and "Is associated with clear metrics and goals so that progress and completion can be identified" received intermediate mean scores of 3.02 and 2.97 respectively and did not differ significantly from each other.

Table 4
Extent to Which Challenges Fulfill Characteristics of Grand Challenges

Challenge	Hard but Doable	Positive Outcomes	Clear Metrics	Popular	Average
Increase Equity	3.12 (.87)	3.54 (.73)	3.04 (.88)	3.23 (.86)	3.20_{a} (.60)
Drive Innovation	3.05 (.78)	3.35 (.75)	2.99 (.96)	2.91 (.86)	3.08_{ab} (.60)
Change Over Time	3.16 (.83)	3.27 (.79)	3.01 (.93)	2.77 (.95)	3.05_{abc} (.66)
Massive Data	3.23 (.83)	3.28 (.77)	2.93 (.90)	2.93 (.88)	3.04_{bc} (.63)
Immediate Improvements	3.09 (.89)	3.40 (.74)	3.07 (.81)	2.60 (.97)	3.01_{bc} (.59)
Disaggregate Data	2.82 (1.02)	3.21 (.81)	3.10 (.89)	2.80 (.86)	2.95_{c} (.63)
Student Self-Evaluation	2.97 (.87)	3.23 (.85)	2.92 (.88)	2.58 (.94)	2.94_{c} (.60)
Communicate	3.01 (.92)	3.09 (.77)	2.85 (.88)	2.80 (.93)	2.93_{c} (.63)
Inform Budget	2.91 (.93)	3.04 (.80)	2.95 (.91)	2.61 (.94)	2.90_{cd} (.62)
ePortfolios	2.84 (1.01)	2.88 (.85)	2.83 (.95)	2.42 (.93)	2.73_{d} (.71)
Average	3.02_{e} (.04)	3.23 (.04)	2.97_{e} (.05)	2.77 (.05)	

Note: Challenges are listed in order of overall score with standard deviations reported in parentheses. Means that do not share subscripts differ at $p < .05$ with the Bonferroni correction.

Rankings of Grand Challenges

In addition to rating challenges using the four characteristics, participants ranked all 10 challenges in order of importance. Table 5 reports the percentage of participants who ranked each challenge as either the top challenge or among the top three. Challenges are ordered from the most to least frequently listed in the top three positions. Table 5 demonstrates a similar pattern as seen in the average scores: Challenges "Increase Equity," "Drive Innovation," and "Change Over Time" were frequently considered among the top three challenges. However, unlike the results of average scores, "Immediate Improvements," appeared frequently among the top challenges and "Massive Data" was rarely listed among the top challenges.

To determine whether the differences in rankings were significant, a Friedman test was calculated on rankings which indicated a significant difference, $\chi^2(9, n = 212) = 246.55$, $p < .001$. Average and median rankings for each challenge are reported in Table 6. Inspection of average rankings revealed four gaps of greater than .60 between challenges "Increase Equity,"

Table 5
Percentage of participants selecting each challenge among the most important challenges

Challenge	% listed in top 3 challenges	% listed in top challenge
Increase Equity	51	21
Immediate Improvement	45	16
Drive Innovation	43	16
Change Over Time	35	8
Student Self-evaluation	32	16
Communicate	28	11
Inform Budget	26	8
Disaggregate Data	16	1
ePortfolios	12	1
Massive Data	11	2

Table 6
Average and Median Rankings of Challenges

Challenge	Median ranking	Average ranking	Standard deviation
Increase Equity	3	4.23	2.87
Drive Innovation	4	4.43	2.66
Immediate Improvement	4	4.21	2.53
Change Over Time	5	4.91	2.43
Communicate	5	5.15	3.00
Student Self-evaluation	5	5.32	2.98
Inform Budget	6	6.83	2.73
Disaggregate Data	6	6.92	2.33
Massive Data	7	7.49	2.69
ePortfolios	8	8.48	2.66

Note: Challenges are listed in order of average ranking. Line breaks indicate significant differences between Challenges.

"Change Over Time," "Inform Budget," "Massive Data," and "ePortfolios." Accordingly, four Wilcoxon Signed-Rank Tests were calculated, comparing challenge "Increase Equity" to "Change Over Time," challenge "Change Over Time" to "Inform Budget," challenge "Inform Budget" to "Massive Data," and challenge "Massive Data" to "ePortfolios." Using a Bonferonni adjusted alpha value of .0125 no significant difference was found between challenge "Increase Equity" and "Change Over Time" (z = 2.40, p = .02, ns) or between challenge "Massive Data" and "ePortfolios" (z = 1.81, p = .07, ns). However, significant differences were observed between "Change Over Time" and "Inform Budget," z = 3.72, $p < .001$ and between challenge "Inform Budget" and "Massive Data," z = 4.55, $p < .001$.

Additional Challenges

In addition to collecting ratings of the 10 challenges, we provided space for participants to propose a challenge. We coded these responses using the grounded theory approach (Charmaz, 2014). This approach includes two phases of coding. During the first phase, narrative data is labeled and categorized according to themes. During the second phase, the labeled categories are reviewed and finalized. The first author grouped the suggestions into categories based on similarity of responses. In instances in which a response included more than one theme, it was divided into separate phrases to group each theme with similar responses. The second author then reviewed the groupings. There were no disagreements in the coding of these responses. A total of 135 suggestions were received (42% of total sample), which were classified into six broad themes. Sample qualitative responses for each theme are reported in Table 7.

As shown in Table 7, the first and most common theme was the need to improve the culture of assessment. This theme was expressed by 66 participants (29% of the total sample) and included suggestions regarding the need to increase buy-in, reduce fear or

The first and most common theme was the need to improve the culture of assessment. This theme was expressed by 66 participants (29% of the total sample) and included suggestions regarding the need to increase buy-in, reduce fear or negativity, integrate assessment with teaching, and engage groups of stakeholders.

Table 7
Sample Challenges Proposed by Participants

Theme	Sample response
Improve Culture of Assessment	Increase buy in. There are too many people who don't see the value. Build assessment into the ongoing, regular routines of higher education. See assessment as an important part of effective teaching and learning.
Improve Measurement of Learning	Determine ways to measure and encourage deep student learning. Develop tools to evaluate learning that are meaningful and actionable. Demonstrate student learning that occurs outside of the classroom.
Use Findings to Improve Learning	Shift focus from box checking towards learning improvement. Use results to improve educational programs. Consistently closing the loop with assessment findings.
Assess Learning Over Time and Across Institutions	Develop valid and reliable assessments for use by multiple institutions. Measure achievement across courses, majors, institutions and over time. Track far-transfer and longitudinal learning.
Increase Resources for Assessment	Financial and human investment in assessment activities. Make assessment less expensive (money and human resources).
Increase Equity for Specific Groups	Use data to remove systemic barriers for marginalized groups. Make race, gender, and SES non-predictive of STEM persistence.

negativity, integrate assessment with teaching, and engage groups of stakeholders. The second theme, which was expressed by 28 participants (12% of total sample), was to improve the measurement of student learning. Participants mentioned the need to consider adopting standard forms of measurement, making comparisons across institutions, and improving the validity and reliability of measures. The third theme, which was expressed by 14 participants (6% of the total sample), was to increase the use of assessment findings to improve student learning. The fourth theme, which was expressed by 13 participants (6% of the total sample), was to increase the assessment of learning over time and across institutions. Several of these suggestions included elements that were similar to "Change Over Time" but included references to specific long-term outcomes of interest, such as graduation rates and employment outcomes. Finally, seven responses (3% of the total sample) indicated a need for increased financial resources to support assessment or less expensive means of assessing student learning, and six responses (3% of the total sample) referred to the need to increase equity for specific underserved groups.

Assessment professionals were most confident that addressing assessment challenges could produce positive outcomes potentially affecting large numbers of people. However, assessment professionals were less confident that the challenges would capture the popular imagination, and thus garner political support.

Discussion

This study was designed to identify the most pressing grand challenges facing the field of assessment in higher education. A review of recent literature revealed active consideration of 10 important challenges facing the field, which were ranked and evaluated by assessment professionals using four characteristics of grand challenges (Singer-Freeman & Robinson, 2020). Some interesting differences emerged regarding how assessment professionals viewed the characteristics of grand challenges, as they relate to assessment. Assessment professionals were most confident that addressing assessment challenges could "produce positive outcomes potentially affecting large numbers of people." However, assessment professionals were less confident that the challenges would "capture the popular imagination, and thus garner political support." Although the challenge "Communicate" did not emerge as a highly endorsed challenge, limited confidence that assessment challenges will "capture the popular imagination, and thus garner political support," may indicate that there is a need to improve communication about the benefits of assessment with individuals outside of higher education.

To learn more about assessment professionals' beliefs about how the characteristics of grand challenges should be weighted, audience members at a national assessment conference presentation completed a brief survey in which they reported whether the four characteristics

should be given equal weight, and if not weighed equally, describe how they should be ranked (prior to the presentation of survey results). Among the 16 attendees who provided feedback, all reported that the characteristics should not be given equal weight. There was a strong consensus that "produce positive outcomes potentially affecting large numbers of people" and "associated with clear metrics and goals so progress and completion can be identified" were more important than being "extremely hard to do, yet doable" or "capture popular imagination, and thus garner political support." Thus, from this small sample, it appears that the characteristics survey respondents felt most fully described the challenges were also the characteristics assessment professionals believe to be the most important.

We view existing educational equity gaps as resulting from failures of practice and are exploring ways assessment and assignment choices can support increased equity in higher education

The survey results identified four challenges that have strong support from assessment professionals. "Use assessment findings to increase equity," which was listed as a top challenge by 51% of respondents, had the highest overall average score across the four characteristics and the highest median rank. As we work to address this challenge in the Grand Challenges in Assessment Project, we are investigating effects of current practices on underserved groups. We view existing educational equity gaps as resulting from failures of practice and are exploring ways assessment and assignment choices can support increased equity in higher education (Blaich & Wise, 2018, Malcom-Piqueux, 2018; Montenegro & Jankowski, 2017; 2020; Singer-Freeman & Bastone, 2019; Singer-Freeman et al., 2019).

The challenge "Produce visible and actionable assessment findings that drive innovation," was listed as a top challenge by 43% of respondents and did not differ from the "Increase Equity" challenge in overall average score or rank. This challenge is related to other highly rated challenges. Successful innovations might increase equity or support rapid improvements in pedagogy. As we work to address this challenge in the Grand Challenges in Assessment Project, we are seeking ways to improve assessment methodology so that we gather evidence that informs our understanding of the outcomes associated with innovative practices. We are reviewing strategies that engage faculty partners to identify causes of gaps in student learning, identify evidence-based solutions, determine whether selected interventions are implemented with high fidelity, and measure the extent to which the interventions drive learning improvements (Eubanks, 2017; Fulcher et al., 2017; Smith et al., 2017; Stevenson et al., 2017; Stitt-Bergh et al., 2018).

The challenge "Examine changes in institutional effectiveness (including student learning) over time" was listed as a top challenge by 35% of respondents and did not differ from the "Increase Equity" challenge in overall average score or rank. As we work to address this challenge in the Grand Challenges in Assessment Project, we are reviewing strategies to improve measurement and tracking of individual students' learning (Baer, 2017; Eubanks, 2019; Miller, 2016; Pasquerella, 2018) as well as progress toward broad institutional goals. To track learning over time effectively, we must find better sources of longitudinal student data. However, there is also a tension between the need for longitudinal data and the need to make rapid changes in instruction or services to support student success. To resolve this tension, it will be important to identify broad metrics that allow the accurate tracking of progress toward goals in a constantly shifting educational landscape.

Finally, the challenge "Use assessment findings to direct immediate pedagogical improvements" was listed as a top challenge by 45% of respondents and did not differ from the "Increase Equity" challenge in median ranking; however, it received a lower overall score than the "Increase Equity" challenge. As we work to address this challenge in the Grand Challenges in Assessment Project, we are seeking to identify new models and methods of assessment and accountability that use relevant findings to make immediate pedagogical changes (Eubanks, 2017; Maki, 2017). To identify effective models, we are evaluating practices and technologies from a range of disciplines, considering socio-cognitive factors that influence student learning, and seeking effective measures of student learning over time (Eynon & Gambino, 2017; López-Pastor & Sicilia-Camacho 2017).

Intermediate levels of support were observed for challenges "Use assessment findings to inform budgetary initiatives," "Involve students in authentic self-evaluation of their own learning," and "Communicate relevant, timely and contextualized information about student learning to stakeholders." These challenges were ranked as among the top three

challenges for only 26 to 32% of respondents, and their overall scores across characteristics and rankings were significantly lower than those for the challenge "Increase Equity." Finally, the lowest levels of support were observed for challenges "Disaggregate data to include important student characteristics," "Use ePortfolios to capture students' learning over the entire span of their education," and "Leverage technology to analyze massive data sets within and across institutions." These challenges were only ranked as among the top three challenges by 11 to 16% of respondents. Interestingly, each of these challenges describe a mechanism by which other, more highly-rated challenges might be achieved. For instance, data disaggregation is an important tool employed to increase equity, and the analysis of massive data sets is a tool that can be used to produce actionable assessment findings. Finally, ePortfolios are used to examine changes in learning over time and engage students in self-evaluation of learning. It may be the connected nature of those challenges as enablers of other challenges that led to the lower ratings.

Limitations

The top challenges that emerged from this study provide confirmation that the field of assessment has moved beyond conducting assessment to demonstrate compliance and is ready to fully embrace the use of assessment for improvement.

Although we found clear patterns of support for certain challenges in the current study, our findings are limited by the use of a survey design. Because the challenges were listed without detailed descriptions of the research literature from which they emerged, it is possible that individuals differed in their interpretation of the stated challenges in ways that influenced their rankings and ratings. We were also limited by the relatively small number of responses to our survey. In particular, we lacked adequate representation from Asian and Hispanic or Latinx assessment professionals and professionals employed at private, non-profit institutions of higher education.

Future Directions

The top challenges that emerged from this study provide confirmation that the field of assessment has moved beyond conducting assessment to demonstrate compliance and is ready to fully embrace the use of assessment for improvement. The participants in this study wish to increase equity, drive innovation, improve pedagogy, and measure progress over time. The identification of grand challenges is only a starting point. For grand challenges to increase the speed of progress in the field of assessment they must be used to coordinate efforts, strengthen commitment from stakeholders, support communication with the public, and attract funds (Gould, 2010; Stephan et al., 2015; Weiss & Khademian, 2019).

Since the completion of this survey, we have launched the Grand Challenges in Assessment Project to create strategic plans that will coordinate research and practical efforts to address the four challenges with the broadest support. The project has endorsements from nine national organizations. Nearly 100 faculty, staff, and students are collaborating in four working groups to create national strategic plans to address the top four challenges identified in the survey. This work is being overseen and supported by a steering committee with representation from each of the endorsing organizations. The working groups include full-time assessment professionals from offices of assessment, institutional effectiveness, and institutional research, as well as faculty members, students, representatives from professional organizations, and representatives from higher education organizations. There is also representation from all types of institutions of higher education from all accrediting regions. The represented institutions include private, public, religious, historically black colleges and universities, community colleges, liberal arts colleges, and research universities. After fully defining each challenge, working groups researched evidence-based routes to improvement, and are currently creating actionable strategic plans for improvement that can be enacted both nationally and locally.

References

Alberts, B. (2013). Prioritizing science education. *Science*, 340(6130). 249. http://doi:10.1126/science.1239041

Ariovich, L., Bral, C., Gregg, P., Gulliford, M., & Morrow, J.A. (2018). The Assessment Profession in Higher Education: Addressing the Varied Professional Development Needs of Practitioners. Lexington, KY: Association for the Assessment of Learning in Higher Education and Watermark. https://go.watermarkinsights.com/AALHEReport-2018_AALHE-Joint-StudySplit.html

Baer, L. (2017). Connecting the dots: Accountability, assessment, analytics, and accreditation. *Planning for Higher Education, 46*(1), 1-16. https://search.proquest.com/docview/2008824562?accountid=14605&rfr id=info%3Axri%2Fsid%3Aprimo

Banta, T. W., Ewell, P. T., & Cogswell, C. A. (2016). Tracing assessment practice as reflected in *Assessment Update. Occasional Paper 28*, National Institute for Learning Outcomes Assessment, 1-30. https://www.learningoutcomesassessment.org/wp-content/uploads/2019/02/OccasionalPaper28.pdf

Blaich, C., & Wise, K. (2018). Highlights of the 2018 HEDS annual conference. *Assessment Update*. 4-5. https://doi.org/10.1002/au.30151

Charmaz, K. (2014). *Constructing Grounded Theory 2nd Edition*. Sage Publications. https://www.amazon.com/Constructing-Grounded-Introducing-Qualitative-Methods/dp/0857029142

Combs, B., & Rose, T. (2016). *Status of Assessment Staff in Higher Education: An Overview*. Paper presented at the annual Association for the Assessment of Learning in Higher Education Conference, Milwaukee, WI.

Eubanks, D. (2017). A guide for the perplexed. *Intersection*. 4-14. https://cdn.ymaws.com/www.aalhe.org/resource/resmgr/docs/Int/AAHLE_Fall_2017_Intersection.pdf

Eubanks, D. (2019). Reassessing the elephant, part 1. *Assessment Update, 31*: 6-7. http://doi:10.1002/au.30166

Eynon, B. & Gambino, L. M. (2017). *High-Impact ePortfolio Practice: A Catalyst for Student, Faculty, and Institutional Learning*. Stylus. https://styluspub.presswarehouse.com/browse/book/9781620365052/High-Impact-ePortfolio-Practice

Frew, E. J., Fuller, M., & Hobbs, M. (2007). *Assessment Professional Salary Survey*. Personal Communication

Fulcher, K. H., Smith, K. L., Sanchez, E. R. H., Ames, A. J., & Meixner, C. (2017). Return of the pig: Standards for learning improvement. *Research & Practice in Assessment, 11*, 10-40. http://www.rpajournal.com/dev/wp-content/uploads/2017/03/A2.pdf

Fuller, M., Henderson, S., & Bustamante, R. (2015). Assessment leaders' perspectives of institutional cultures of assessment: a Delphi study, *Assessment & Evaluation in Higher Education*, 40:3, 331-351. 10.1080/02602938.2014.917369

Gould, M. (2010). GIScience grand challenges: How can research and technology in this field address big-picture problems? *ArcUser, 13(4)*, 64–65. http://www.esri.com/news/arcuser/1010/files/geochallenges.pdf

Harpe, S. E. (2015). How to analyze Likert and other rating scale data. *Currents in Pharmacy Teaching and Learning, 7*, 836-850. http://dx.doi.org/10.1016/j.cptl.2015.08.001

Hilbert, D. (1902). Mathematical problems. *Bulletin of the American Mathematical Society, 8*(10), 437–479. http://doi:10.1090/S0002-9904-1902-00923-3

Jaschik, S., & Lederman, D. (2020). 2020 Survey of College and University Chief Academic Officers, *Inside Higher Education*, https://www.watermarkinsights.com/resources/ebooks-guides/inside-higher-ed-2020-survey-of-college-and-university-chief-academic-officers

López-Pastor, V., & Sicilia-Camacho, A. (2017). Formative and shared assessment in higher education. Lessons learned and challenges for the future, *Assessment & Evaluation in Higher Education, 42:1*, 77-97. https://doi.org/10.1080/02602938.2015.1083535

Maki, P. L. (2017). *Real-Time Student Assessment: Meeting the Imperative for Improved Time to Degree, Closing the Opportunity Gap, and Assuring Student Competencies for 21st-Century Needs*. Stylus Publishing, LLC. https://www.amazon.com/Real-Time-Student-Assessment-Competencies-21st-Century/dp/1620364883

Malcom-Piqueux, L. (2018). Making sense of data in equity-minded ways. In *A Vision for Equity: Results from AAC&Us Project Committing to Equity and Inclusive Excellence: Campus-Based Strategies for Student Success*. https://www.aacu.org/publications/vision-equity

Miller, T. (2016). *Higher Education Outcomes-Based Funding Models and Academic Quality. Lumina Issue Papers*, 1-17. https://www.luminafoundation.org/files/resources/ensuring-quality-1.pdf

Montenegro, E., & Jankowski, N. A. (2017). *Equity and assessment: Moving towards culturally responsive assessment (Occasional Paper No. 29)*. Urbana: University of Illinois and Indiana University, NILOA. https://www.learningoutcomesassessment.org/documents/OccasionalPaper29.pdf

Montenegro, E., & Jankowski, N. A. (2020). A new decade for assessment: Embedding equity into assessment praxis (Occasional Paper No. 42). Urbana, IL: University of Illinois and Indiana University, NILOA. https://www.learningoutcomesassessment.org/wp-content/uploads/2020/01/A-New-Decade-for-Assessment.pdf

National Academy of Engineering. 2016. *Grand Challenges for Engineering: Imperatives, Prospects, and Priorities*. Washington: National Academies Press. doi: 10.17226/23440

National Research Council (2005). *Sustainability in the Chemical Industry: Grand Challenges and Research Needs—A Workshop Report*. Washington, DC: National Academies Press. http://www.nap.edu/catalog/11437.html

Nicholas, M. C., & Slotnick, R. C. (2018). *A Portrait of the Assessment Professional in the United States: Results from a National Survey*. (Occasional Paper 34) Urbana, IL: University of Illinois and Indiana University, NILOA. http://learningoutcomesassessment.org/occasionalpaperthirtyfour.html

Omenn, G. S. (2006). Grand challenges and great opportunities in science, technology, and public policy. *Science, 314*, 1696-1704. http://science.sciencemag.org/content/sci/314/5806/1696.full.pdf?sid=307f6571-c04c-4721-98e4-591981642fe7

Pasquerella, L. (2018). Planning for America's future: Educating for democracy. *Liberal Education, 104*(1). https://www.aacu.org/liberaleducation/2018/winter/pasquerella

Popowitz, M., & Dorgelo, C. (2018). Report on University-Led Grand Challenges. *UCLA: Grand Challenges*. https://escholarship.org/uc/item/46f121cr

Singer-Freeman, K. E., & Bastone, L. (2019). Increasing equity in general education through self-relevant writing. *Intersection*, 24-27. https://cdn.ymaws.com/www.aalhe.org/resource/resmgr/AAHLE_fall_2019_Intersection.pdf

Singer-Freeman, K. E., Hobbs, H., & Robinson, C. (2019). Theoretical matrix of culturally relevant assessment. *Assessment Update, 31*(4). https://doi.org/10.1002/au.30176

Singer-Freeman, K. E., & Robinson, C. (2020). Grand challenges in assessment: Collective issues in need of solutions. *NILOA Occasional Paper 49*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment. https://www.learningoutcomesassessment.org/wp-content/uploads/2020/11/GrandChallenges.pdf

Smith, K. L., Finney, S. J., & Fulcher, K. (2017). Actionable steps for engaging assessment practitioners and faculty in implementation fidelity research. *Research and Practice in Assessment, 12*, 71-86. http://www.rpajournal.com/actionable-steps-for-engaging-assessment-practitioners-and-faculty-in-implementation-fidelity-research/

Stephan, M. L., Chval, K. B., Wanko, J. J., Civil, M., Fish, M. C., Herbel-Eisenman, B., Konold, C., & Wilkerson, T. L. (2015). Grand challenges and opportunities in mathematics education research. *Journal for Research in Mathematics Education, 46*(2). https://www.nctm.org/Publications/journal-for-research-in-mathematics-education/2015/Vol46/Issue2/Grand-Challenges-and-Opportunities-in-Mathematics-Education-Research/

Stevenson, J. F., Finan, E., & Martel, M. (2017). Measuring assessment climate: A developmental perspective. *Research and Practice in Assessment, 12*, 41-57. https://files.eric.ed.gov/fulltext/EJ1168688.pdf

Stitt-Bergh, M., Kinzie, J., & Fulcher, K. (2018). Refining an approach to assessment for learning improvement. *Research and Practice in Assessment, 13*, 27-33. http://www.rpajournal.com/refining-an-approach-to-assessment-for-learning-improvement/

Uehara, E. S., Barth, R. P., Olson, S., Catalano, R. F., Hawkins, J. D., Kemp, S., Nurius, P. S., Padgett, D. K., & Sherraden, M. (2014). *Identifying and tackling grand challenges for social work* (Grand Challenges for Social Work Initiative Working Paper No. 3). Baltimore, MD: American Academy of Social Work and Social Welfare. https://aaswsw.org/wp-content/uploads/2015/12/WP3-with-cover.pdf

Varmus, H., Klausner, R., Zerhouni, E., Acharya, T., Daar, A. S., & Singer, P. A. (2003). Grand challenges in global health. *Science, 302*(5644), 398–399. http://doi:10.1126/science.1091769

Weiss, J. A., & Khademian, A. (2019). An analysis of the pros and cons of universities' grand challenges (opinion). *Inside Higher Education*. https://www.insidehighered.com/print/views/2019/09/03/analysis-pros-and-cons-universities-grand-challenges-opinion

Appendix

Grand Challenges Survery

Dear Colleague,

I hope you will take a few minutes to consider helping us to identify and prioritize grand challenges facing assessment professionals. A number of national and international organizations have compiled lists of grand challenges in their fields. For example, see Omenn's (2006) discussion of great challenges in Science. The identification of grand challenges can be a useful process that unifies the efforts of practitioners in a field. Unified efforts increase the possibility of creating meaningful and lasting progress. For the purposes of our work we modified the set of characteristics for grand challenges proposed by Gould (2010) and Stephan et al. (2015) resulting in the following characteristics of grand challenges:

(1) Extremely hard to do, yet doable;

(2) Produce positive outcomes potentially affecting large numbers of people;

(3) Associated with clear metrics and goals so that progress and completion can be identified

(4) Capture the popular imagination, and thus garner political support.

Do you believe there should be any additions, deletions, or changes to the characteristics of grand challenges described above? If so, please share your suggestions here.

For each of the following goals, please indicate the extent to which you believe it fulfills the four characteristics of grand challenges.

Involve students in authentic self-evaluation of their own learning.

For each of the following goals, please indicate the extent to which you believe it fulfills the four characteristics of grand challenges.

Involve students in authentic self-evaluation of their own learning.

	Not at all	A little	To some extent	Very much
1) Extremely hard to do, yet doable.	○	◉	○	○
2) Would produce positive outcomes potentially affecting large numbers of people.	○	○	○	○
3) Is associated with clear metrics and goals so that progress and completion can be identified.	○	○	○	○
4) Would capture the popular imagination, and thus garner political support.	○	○	○	○

Use assessment findings to increase equity.

	Not at all	A little	To some extent	Very much
1) Extremely hard to do, yet doable.	○	○	○	○
2) Would produce positive outcomes potentially affecting large numbers of people.	○	○	○	○
3) Is associated with clear metrics and goals so that progress and completion can be identified.	○	○	○	○
4) Would capture the popular imagination, and thus garner political support.	○	○	○	○

Produce visible and actionable assessment findings that drive innovation.

	Not at all	A little	To some extent	Very much
1) Extremely hard to do, yet doable.	○	○	○	○
2) Would produce positive outcomes potentially affecting large numbers of people.	○	○	○	○
3) Is associated with clear metrics and goals so that progress and completion can be identified.	○	○	○	○
4) Would capture the popular imagination, and thus garner political support.	○	○	○	○

Use ePortfolios to capture students' learning over the entire span of their education.

	Not at all	A little	To some extent	Very much
1) Extremely hard to do, yet doable.	○	○	○	○
2) Would produce positive outcomes potentially affecting large numbers of people.	○	○	○	○

	Not at all	A little	To some extent	Very much
3) Is associated with clear metrics and goals so that progress and completion can be identified.	○	○	○	○
4) Would capture the popular imagination, and thus garner political support.	○	○	○	○

Examine changes in institutional effectiveness (including student learning) over time.

	Not at all	A little	To some extent	Very much
1) Extremely hard to do, yet doable.	○	○	○	○
2) Would produce positive outcomes potentially affecting large numbers of people.	○	○	○	○
3) Is associated with clear metrics and goals so that progress and completion can be identified.	○	○	○	○
4) Would capture the popular imagination, and thus garner political support.	○	○	○	

Use assessment findings to direct immediate pedagogical improvements.

	Not at all	A little	To some extent	Very much
1) Extremely hard to do, yet doable.	○	○	○	○
2) Would produce positive outcomes potentially affecting large numbers of people.	○	○	○	○
3) Is associated with clear metrics and goals so that progress and completion can be identified.	○	○	○	○
4) Would capture the popular imagination, and	○	○	○	○

	Not at all	A little	To some extent	Very much
thus garner political support.				

Use assessment findings to inform budgetary initiatives.

	Not at all	A little	To some extent	Very much
1) Extremely hard to do, yet doable.	○	○	○	○
2) Would produce positive outcomes potentially affecting large numbers of people.	○	○	○	○
3) Is associated with clear metrics and goals so that progress and completion can be identified.	○	○	○	○
4) Would capture the popular imagination, and thus garner political support.	○	○		

Disaggregate data to consider important student characteristics.

	Not at all	A little	To some extent	Very much
1) Extremely hard to do, yet doable.	○	○	○	○
2) Would produce positive outcomes potentially affecting large numbers of people.	○	○	○	○
3) Is associated with clear metrics and goals so that progress and completion can be identified.	○	○	○	○
4) Would capture the popular imagination, and thus garner political support.	○	○	○	○

Leverage technology to analyze massive data sets within and across institutions.

	Not at all	A little	To some extent	Very much
1) Extremely hard to do, yet doable.	○	○	○	○
2) Would produce positive outcomes potentially affecting large numbers of people.	○	○	○	○
3) Is associated with clear metrics and goals so that progress and completion can be identified.	○	○	○	○
4) Would capture the popular imagination, and thus garner political support.	○	○	○	

Communicate relevant, timely, and contextualized information about student learning to stakeholders.

	Not at all	A little	To some extent	Very much
1) Extremely hard to do, yet doable.	○	○	○	○
2) Would produce positive outcomes potentially affecting large numbers of people.	○	○	○	○
3) Is associated with clear metrics and goals so that progress and completion can be identified.	○	○	○	○
4) Would capture the popular imagination, and thus garner political support.	○	○	○	○

What would you propose as a grand challenge for assessment in higher education?

Please rate the challenge you proposed using the four characteristics of grand challenges.

	Not at all	A little	To some extent	Very much
1) Extremely hard to do, yet doable.	○	○	○	○

	Not at all	A little	To some extent	Very much
2) Would produce positive outcomes potentially affecting large numbers of people.				
3) Is associated with clear metrics and goals so that progress and completion can be identified.				
4) Would capture the popular imagination, and thus garner political support.				

Please rank order these challenges from least to most important. If you did not propose a grand challenge please select "least important" for line #11.

	Least important	1	2	3	4	5	6	7	8	9	Most Important
1) Involve students in authentic self-evaluation of their own learning.											
2) Use assessment findings to increase equity.											
3) Produce visible and actionable assessment findings that drive innovation.											
4) Use ePortfolios to capture students' learning over the entire span of their education.											
5) Examine changes in student learning and institutional effectiveness over time.											
6) Use assessment findings to direct immediate pedagogical improvements.											

	Least important	1	2	3	4	5	6	7	8	9	Most Important
7) Use assessment findings to inform budgetary initiatives.	○	○	○	○	○	○	○	○	○	○	○
8) Disaggregate data to consider important student characteristics.	○	○	○	○	○	○	○	○	○	○	○
9) Leverage technology to analyze massive data sets within and across institutions.	○	○	○	○	○	○	○	○	○	○	○
10) Communicate relevant, timely and contextualized information about student learning to stakeholders.	○	○	○	○	○	○	○	○	○	○	○
11) The Grand Challenge you proposed above.	○	○	○	○	○	○	○	○	○	○	○

How many years have you been employed in higher education?

How many years have you been involved in assessment activities?

Which of the following best describes your current institution:

- ☐ Two-Year Institution
- ☐ Four-Year Primarily Undergraduate Institution
- ☐ Undergraduate and Graduate Institution
- ☐ Primarily Graduate Institution
- ☐ Other (please explain)

Which of the following best describes your current institution?

- ○ Public
- ○ Private Non-profit
- ○ Private For-profit

Which of the following describes your current position?

- ☐ Higher education administration
- ☐ Higher education teaching
- ☐ Higher education research
- ☐ Professional organization serving higher education
- ☐ Accrediting organization
- ☐ Other (please explain)

Sex

- ○ Female
- ○ Male
- ○ Another

Race and Ethnicity (please select all that apply)

- ☐ Alaska Native
- ☐ American Indian
- ☐ Asian
- ☐ Black or African American
- ☐ Hispanic or Latino
- ☐ Native Hawaiian
- ☐ Other Pacific Islander
- ☐ White

Age

Thank you for completing our survey!

If you would like to be invited to future conversations about the grand challenges facing assessment practitioners or receive information about the results of this survey, please provide contact information below.

Please be sure to click on "finish" below so your answers will be submitted.

Abstract

Prior research suggests that Academic Program Review (APR) is most effective when it is a systematic process that supports program improvement. One potential way to increase faculty involvement in comprehensive APR processes is by engaging faculty members as internal peer reviewers (IPRs). This qualitative study investigated faculty members' experiences as IPRs of academic programs within their home institution but outside of their own departments. Semi-structured interviews were conducted with 14 faculty members at a public, research-extensive university who served as IPRs. Data analysis suggests that faculty members who engaged as IPRs (1) gained a deeper understanding of the APR process, (2) learned more about the work of other departments on campus, and (3) learned how to more effectively discuss and engage in assessment within their home departments. Suggestions are provided for structuring APR processes in ways that may further develop and promote a positive culture of assessment.

AUTHORS

Hannah P. Davis, M.Ed.
Virginia Tech

Katherine S. Biddle, Ph.D., LPC
REAT
Carilion Clinic

Molly R. Hall, Ph.D.
Virginia Tech

Academic Program Review: Examining the Experiences of Faculty Members Serving as Internal Peer Reviewers

In addition to teaching classes and conducting research, faculty contribute their time and energy to a wide variety of campus programs and committees in an effort to enhance student learning and the educational environment. Many want these experiences to be impactful rather than time spent "spinning their wheels." Perhaps they, like some of the participants in this study, are looking to participate in meaningful experiences that they can both contribute to and gain value from. This study suggests that utilizing faculty as Internal Peer Reviewers (IPRs) during Academic Program Review (APR) may provide an opportunity for faculty to do just that.

Academic programs in higher education are facing increasing demands to provide evidence of educational quality. These demands translate to increased expectations for assessment and accountability. APR is one example of a continuous, systematic process supporting ongoing quality assurance, program improvement, institutional autonomy (Creamer & Janosik, 1999), and an improved ability to advocate for new resources (Banta, 2014). The majority of U.S. higher education institutions began employing some type of APR process during the early 1980s (e.g., Barak, 1982), and APR has since become a common practice in most colleges and universities (Rickards & Stitt-Bergh, 2016). However, of the limited literature that exists regarding best practices and the effectiveness of APR, the majority is theoretical rather than empirical (McGowan, 2019).

CORRESPONDENCE

Email
mrhall@vt.edu

The use of IPRs, faculty members who review programs within their home institution but outside of their own department, is one such rarely examined practice that warrants further consideration. IPRs may benefit the academic programs being reviewed by offering unique perspectives that are not domain specific, encouraging collegiality, and increasing consideration of student learning at the program level (Bloom, 2010). IPRs may also reap professional benefits through developing collaborations across disciplines and participating in a meaningful assessment process centered on student learning and quality improvement (Banta, 2014).

The use of IPRs, faculty members who review programs within their home institution but outside of their own department, is one such rarely examined practice that warrants further consideration.

In this article we present a qualitative study conducted to explore the experiences of faculty members who participate as IPRs of other academic programs. This research builds on the limited literature concerning best practices and the importance of faculty involvement in APR. We provide a description of the APR process, report major findings, and examine how an internal peer review process can be used to further develop and promote a positive culture of assessment. The results fill an important gap in the literature on APR practices by elucidating the benefits of employing internal peer review as one element of a comprehensive APR process. This article is intended to be useful to individuals and institutions attempting to increase faculty involvement in assessment activities.

Academic Program Review

Academic programs are held accountable through a number of review processes, including regional and/or professional accreditation, student learning outcomes assessment, performance-based budgeting, and formal program review. Even for those departments not required to participate in formal reviews by an official accrediting agency, there are often external pressures from administrators to demonstrate evidence of ongoing program improvement (Colón & Dana, 2015). At the onset of the assessment movement in the 1980s, periodic program reviews focused primarily on the availability of resources to operate a program and included few, if any, indicators of performance that might lead to meaningful programmatic change (Bresciani, 2006; Gentemann et al., 1994).

A comprehensive APR process should be tailored to the individual institution and program being evaluated, include student learning outcome data, and be undertaken as an ongoing process focused on program improvement. Ideally, APR should lead to informed decision-making regarding curricula and student learning (Rodgers et al., 2013). Given increasing demands for high quality education, calls to center student learning as the primary focus of program review have been prominent since at least the early 1990s (e.g., Gentemann et al., 1994). Yet, in a survey of 130 institutions across Carnegie categories, Wergin and Swingen (2000) determined that, with few exceptions, by 2000 most institutions still did not include student learning outcome data in the departmental evaluation process. This had changed by the time the National Institute of Learning Outcomes Assessment conducted a national survey in 2013, and recent literature now suggests that the majority of program review processes do incorporate student learning outcomes (e.g., McGowan, 2019). However, in some cases, outcomes are still reviewed for the sole purpose of meeting accreditation standards rather than as part of an effort to make meaningful changes to courses or programs (Blumberg, 2017; Kuh et al., 2014).

Faculty Involvement in Assessment

Effective facilitation of a meaningful APR process necessitates the active involvement of faculty members (Maki, 2004). In fact, systematic program review processes can be used as a tool to proactively involve faculty members in program decision-making (Shambaugh, 2017). As experts in their own program(s) and participants in the unique culture of their institution, faculty are in a position to be deeply aware of programmatic needs. Additionally, many faculty members are engaged in activities centered on student growth and achievement and possess an innate intellectual curiosity about their students' learning (Maki, 2004). Faculty commitment to assessment is crucial to focusing curriculum on student learning, fostering positive programmatic changes, and promoting a positive culture of assessment (Ndoye & Parker, 2010).

Though faculty involvement is critical to the facilitation of a meaningful APR process, faculty members have not always viewed the process as useful. At the turn of this century, Wergin and Swingen (2000) found that most faculty members did not identify the APR process as positively affecting their professional practice. Given a perceived lack of impact on programs and poor integration as a systematic practice within an institution, many characterized APR as burdensome and ritualistic. Resistance persists today, as many faculty members perceive the review process as authoritarian and non-collegial (Bowker, 2016). Faculty members are more likely to embrace APR undertaken for the purpose of program improvement (Novodvorsky et al., 2015; Rodgers et al., 2013; Townley et al., 2003). If faculty members are unaware of the quality improvement focus of the review, they may not recognize its utility or participate meaningfully in the process (Bresciani, 2006; Rodrigues, 2002; Wergin, 1999). Conversely, faculty who view assessment activities as being improvement driven and/or meaningful are more likely to acknowledge their value and embrace future assessment practices (Rodgers et al., 2013; Trullen & Rodríguez, 2013).

Emil and Cress (2014) indicate faculty attitudes and beliefs towards assessment affect faculty members' willingness to engage in assessment activities. For many years, increased demands for accountability have amplified reservations about assessment in general (Gentemann et al., 1994), and faculty resistance to assessment practices is well documented (e.g., Bowker, 2016; Rodgers et al., 2013; Shavelson, 2010). For example, in a survey of faculty members from business programs, Pringle and Michel (2007) found that of the 43% who acknowledged resistance to assessment practices, more than half felt overwhelmed or overloaded by assessment-related activities. Other frequently cited reasons for faculty resistance to assessment include fear of evaluation, loss of academic freedom, and little return on investment (Linkon, 2005).

Appropriate strategies are required not only to increase faculty engagement in assessment, but also to bridge the gap between APR best practices and actual review processes. Ideally, these strategies will also aid in the creation of program assessment processes that focus primarily on institutional and student learning improvements (Rickards & Stitt-Bergh, 2016). This suggests a need to address and improve faculty perceptions of and participation in assessment activities. One strategy for improving faculty perceptions and participation is to engage faculty as IPRs in the APR process.

Internal Peer Review Teams

When an internal peer review is conducted as part of an APR process that assesses student learning and is focused on program improvement, faculty participation may result in a more meaningful process, dissemination of best practices, and improvement of the overall culture of assessment within an institution (Ketunnen, 2010). For IPRs, enhanced understanding of the purpose and value of APR may deepen and positively influence change within reviewers' own programs.

As individual faculty members begin to view assessment practices more positively, they can share these experiences with others both within and outside of their own programs, thereby contributing to a positive culture of assessment within their institution.

Cross-evaluation is a procedure in which representatives from various areas of an institution come together to evaluate a designated program through constructive communication to promote and encourage learning and to disseminate best practices of assessment within the institution (Ketunnen, 2010). Internal reviewers gain firsthand knowledge about the workings of other programs, including similarities to and differences from their own campus unit (Banta, 2014). Constructive communication between faculty members from different departments promotes innovation and learning from diverse experiences and views (Ketunnen, 2010). As individual faculty members begin to view assessment practices more positively, they can share these experiences with others both within and outside of their own programs, thereby contributing to a positive culture of assessment within their institution.

Similarly, by sharing their experiences and assessment knowledge with others in their programs and institutions, IPRs may aid in disseminating best practices. Many faculty members do not receive formal training in assessment and are not always aware of resources or other assistance available to them. Those faculty who participate in internal peer review

have greater opportunities to learn assessment skills, gain experience, and draw on available resources, including assessment professionals.

Incorporating Internal Peer Review Teams at Virginia Tech

Virginia Tech's current APR process was launched in 2015. Prior to 2015, using IPRs for APR was not a common practice. The APR process serves as a mechanism for ongoing, systematic review of academic departments and programs with the explicit purpose of fostering continuous improvement. Each academic department conducts a comprehensive evaluation of its activities every five to six years. The process emphasizes reflection, conversation, and feedback in order to facilitate a strong vision for the future through an honest assessment of program strengths, weaknesses, and opportunities for improvement. While Virginia Tech utilizes the typical APR stages of self-study, review, and final report (DiBiasio & Ecker, 1982), the process is supplemented by continuous administrative support from the Institutional Effectiveness (IE) unit; for example, an IE professional sits on each peer review team.

Peer review teams are a mix of individuals with disciplinary expertise similar to the unit being reviewed and expertise distinct from the unit in order to provide diverse perspectives. When APR was first launched in 2015, only IPRs were utilized, but now departments may choose either a team of all IPRs or a mix of internal and external reviewers. While several departments have utilized external reviewers, most departments participating in APR have opted for a review team that consists solely of IPRs. Although departments may nominate reviewers to serve on their peer review team, at least one member of the team is selected from a pool of Virginia Tech faculty members and administrators who have expressed interest in serving as an IPR. Departmental peer review teams are finalized in consultation with the department chair/program director. Both internal and external peer reviewers are offered a $500 stipend per completed review and IPRs may serve on a maximum of two review teams per academic year. Once a peer review team is finalized, all of the reviewers on the team participate in an initial training meeting led by IE professionals at which the APR process, timeline, and peer reviewer responsibilities are discussed in detail.

Steps in the APR Process

By sharing their experiences and assessment knowledge with others in their programs and institutions, IPRs may aid in disseminating best practices.

Each participating department/program completes a self-study report (SSR) designed to encourage departments to reflect on their current operations, develop a vision for the future, and create and implement a plan for continuous improvement. The analysis included in the report is informed by data provided to the department from Institutional Research and other sources, as well as faculty interests and current trends in the field. The SSR is submitted to IE professionals who distribute it to members of the peer review team. Peer reviewers independently analyze the SSR using a rubric designed by IE. An IE professional then facilitates a team meeting to discuss the SSR and identify questions that the review team would like to ask the department. The peer review team then participates in a face-to-face conversation with departmental representatives to discuss the SSR, where the department hopes to be in the future, and how the department plans to get there.

The IE professional sitting on the review team is responsible for compiling the team members' individual rubrics and creating a draft of the review team's report to the department/program. The APR rubric serves as the template for the peer review team's report. When the draft report is complete, the IE professional sends it to the rest of the peer review team to review and edit, and the entire team works together to finalize the report. After the review team's report is finalized and shared with the department/program, face-to-face conversations between each department and its respective dean(s) are held to discuss program review findings and, most importantly, the department's plans for moving forward. These conversations provide an opportunity to discuss implementation items and resource priorities. After each face-to-face conversation with a department, the respective dean writes a memo noting their conversation with the department, the department's plans for moving forward, and the dean's expectations for what the program will accomplish by the time of the next scheduled APR. This memo is shared with the department and the IE unit, which is responsible for archiving all APR materials.

Methods

Research Design

The interpretive framework underlying this research is constructivism, which acknowledges multiple viewpoints and realities and assumes that individuals play an active role in making meaning from their experiences and interactions with others (Creswell, 2013; Jones et al., 2014). Given the gaps in the existing literature and the need for rich, in-depth data related to how faculty perceive their experiences participating as IPRs in APR, the research design is a basic qualitative study (Merriam, 2009). As Patton (2002) asserts "[T]here is a very practical side to qualitative methods that simply involves asking open-ended questions of people and observing matters of interest in real-world settings in order to solve problems, improve programs, or develop policies (pp. 135-136)."

Participants

18 faculty members who served as IPRs.... participate[d] in the study. ...The participants worked in a wide variety of academic disciplines including the arts, humanities, life sciences, and social sciences, as well as professional disciplines such as business, education, and engineering.

After obtaining Institutional Review Board approval, 18 faculty members who served as IPRs during the first two review cycles (Fall 2016 and Fall 2017) of Virginia Tech's relaunched APR process were invited to participate in the study by the third author. Of these, 14 (78%) completed interviews. Table 1 displays basic demographic information about these participants. All study participants were full-time employees at Virginia Tech. The participants worked in a wide variety of academic disciplines including the arts, humanities, life sciences, and social sciences, as well as professional disciplines such as business, education, and engineering. The total number of years that faculty participants had worked in higher education ranged from a minimum of five years to a maximum of 40 years with a mean of 22 years of experience. Participants from two different APR cycles were interviewed to ensure that data were not overly influenced by circumstances occurring during a specific academic year.

Table 1
Characteristics of Study Participants

Characteristic	Frequency
Faculty Rank	
Full Professors	7
Associate Professors	4
Assistant Professors	3
Gender	
Female	8
Male	6
Semester Served as IPR	
Fall 2016	9
Fall 2017	5

Prior Experience as Reviewers

Of these participants, only three had previously participated on an internal APR peer review team; two had prior experience in conducting external APR peer reviews. Nine had no previous formal experience with any type of APR, but six of those respondents discussed participation in peer review in the context of articles, book prospectuses, promotion and tenure, and/or working on departmental committees preparing materials for a review team. Two had been engaged in accreditation activities for either their own department or similar departments at other institutions, while another had significant experience in conducting external reviews on behalf of an accrediting agency.

Data Collection

Data for this study were collected through semi-structured interviews, conducted in two separate rounds of data collection. All interviews were conducted by the first two authors who had no prior interaction with the study participants. The third author did interact with each study participant during the APR process, but did not conduct any of the interviews. Faculty (n = 9) who participated on peer review teams during the Fall 2016 semester were interviewed during Spring 2017, and faculty (n = 5) who participated on peer review teams during the Fall 2017 semester were interviewed during Spring 2018. Each interview was conducted in person at a location selected by the research participant. Interviews were a maximum of 60 minutes in length.

An informed consent form was signed at the beginning of each interview. In addition to being invited to respond to the interview questions (see Appendix A), participants were asked to complete a brief demographic questionnaire (see Appendix B). The second author interviewed participants from the Fall 2016 review cohort, and the first author interviewed participants from the Fall 2017 review cohort. The first author transcribed interviews from both cohorts. Member checks were conducted after transcription; a copy of each interview transcript was sent to the respective participant for review and feedback.

Data Analysis

Transcripts of the semi-structured interviews revealed that responses tended to address both the presented question, as well as previous or unasked questions. Structural codes, codes based on specific elements of the research questions, can be useful when "respondents return to earlier topics or make a cognitive leap" to topics addressed in later questions (MacQueen et al., 1998, p. 33). These structural codes served as an index for applying more focused coding and allowed for the synthesis of data for the purpose of thematic analysis (Auerbach et al., 1998).

This study represents a convergence of information from a variety of IPRs who participated in APR during a two-year period at a single site.

Utilizing the structural codes, the authors were then able to analyze indexed responses and define themes based on participants' experiences. To do so, the constant comparative method (Glaser & Strauss, 1967) was applied. This process involved a thorough analysis of each respondent's comments within their own interview, as well as within the greater context of all of the interviews. The emergent codes from all interviews were defined, compared, and refined until the authors felt they had identified all relevant themes and integrated them into a coherent explanation of the general IPR experience. The qualitative software tool NVivo was used throughout the data analysis process.

Limitations

Acknowledging the limitations of a study is important to provide readers with a deeper understanding of the scope and nature of the research. This study represents a convergence of information from a variety of IPRs who participated in APR during a two-year period at a single site. This single-site study ensured consistency in the expectations and requirements of the institution's specific APR process, but limits the understanding of the broader APR experience as conducted under varying institutional requirements. Additionally, while participants were not offered a monetary incentive for participating in the study, they were provided with a stipend by the IE unit for participating in the APR process. This stipend may have impacted participants' decisions about whether or not to participate in the research study. Potential decreases in discretionary funds might not allow for such stipends to be provided at this or other institutions in the future, in turn impacting which faculty members may choose to participate in assessment-related activities. Finally, while participants and interviewers did not have a prior relationship, the interviewees were invited to participate in the study by the third author who had previously interacted with each participant during the APR process. These prior interactions by the third author may have impacted which faculty members were willing to be interviewed for this study.

Findings and Discussion

Findings from the qualitative analysis of semi-structured interviews show a clear pattern aligning with best practices of APR. Participants identified three beneficial components of the APR process: (1) an organized structure, (2) a focus on improvement, and (3) the use of diverse peer review teams. The support provided by IE professionals emerged as another important aspect of the APR experience. The combination of the APR process and the support received from IE professionals allowed participants to build three avenues of learning as a result of their participation: (1) learning about the purpose and structure of the APR process, (2) learning about other departments, and (3) learning how to engage in and discuss assessment within their home departments. Two minor themes, cross-disciplinary interactions and service to others, are also briefly discussed. Table 2 summarizes the frequencies of qualitative codes from the interview respondents. Additional findings include factors motivating participation, recruitment, and engaging in future service on an internal peer review team. Please note that all quoted remarks come from participant interview data; however, specific quotes are not attributed to individual participants.

Participants identified three beneficial components of the APR process: (1) an organized structure, (2) a focus on improvement, and (3) the use of diverse peer review teams.

Table 2
Qualitative Codes and Frequencies

Code	Frequency
P1: Challenge	17
P2: Effort	8
P3: General	7
P4: Improvement Focused	18
P5: Structure	16
P6: Team Composition	26
S1: Professional Assistance	16
S2: Resources	5
V1: Cross-Disciplinary Interactions	10
V2: Learning (Other Departments)	19
V3: Learning (APR)	9
V4: Learning (Home Departments)	10
V5: Learning (Other)	5
V6: Service to Others	6

Note. Process codes are P1-P6; Support codes are S1-S2; Value codes are V1-V6

Benefits of the APR Process

Participants noted three key elements of the APR process that were beneficial: (1) an organized structure, (2) an improvement focus, and (3) the use of diverse peer review teams. In addition to utilizing the three primary stages of self-study, review, and final report suggested by DiBiasio & Ecker (1982), this APR process included reflection, feedback, and discussion within the structure of the process. Participants felt the structure of the APR process was "well executed" and "streamlined." Review teams felt there was strength "in terms of our engagement, what we were told, how we got oriented, and the things that we were given." One participant stated, "the process isn't, I don't want to say intense, but it was thorough." The structure allowed IPRs "group time, but then also individual time to look over reports and add what it is we think is important." Being given time to work both independently and as a group allowed reviewers to discuss "different perspectives to decide 'what do I really think now that I know that?'" Some participants felt having an APR structure "that gives you that mix, that blend" was ideal.

In addition to the mechanics of the APR process, participants highlighted that it was "really about helping the department." As suggested by Novodvorsky et al. (2015), faculty members seemed to embrace APR because it centered on program improvement. One respondent indicated that in comparison to similar reviews they had participated in,

> This felt like the process was genuine in helping this department get better. And it felt like – it was almost like a fresh set of eyes to help them maybe see some things differently or suggest some things, which I thought was really really positive.

According to Banta (2014) and Flaherty (2016), departments under review reap quality results when they too recognize the APR process as meaningful and focus on improvement. Some reviewers expressed concern about viewing the purpose of program review differently from the department being reviewed, where the latter "took it more as an evaluative [process] and we took it as more of a formative assessment." This was reflected in the way one reviewer felt the department

> presented as a lot of, 'here's what we've done, here's our accomplishments, here are our numbers' and we were really, I think, more interested in where are you going, what are your opportunities, build on your strengths to get to the next level, or to sell yourself within the college in a better way.

Reviewers wanted faculty to be "aware that [APR] is not a punitive thing. We aren't being the curriculum police. We're just trying to find ways to make it better."

Review teams were comprised of people from different departments and colleges, representing different experience levels and research approaches. The practice of selecting a diverse group of IPRs is one way of encouraging the type of constructive communication Ketunnen (2010) and Bloom (2010) have deemed necessary for promoting innovation and learning. One participant stated, "We had different kinds of questions. Different kinds of perspectives, which I think was very useful." While some might be concerned about finding common ground in working in a diverse group, ultimately "the questions [they] were asked to address were general enough that people from…any discipline would be familiar with what was required to do it." This diversity "just worked well. It was a good approach" and "that's a strength to [APR]."

While some might be concerned about finding common ground in working in a diverse group, ultimately "the questions [they] were asked to address were general enough that people from…any discipline would be familiar with what was required to do it."

Respondents indicated that although they put forth significant effort during the process, prospective faculty members might have a "misperception as to the workload." Contrary to the findings of Wergin and Swingen (2000), 93% of IPRs in this study did not characterize the APR process as burdensome. Although participants were informed that the time commitment for participating in APR was approximately 15 hours, only one respondent stated they would not participate again because "Wow! It was a lot of time." Given the level of support from each team's IE professional, IPRs were not required to "fill out a ton of reports." Other participants reported "it was not a huge commitment of time," "it's not painful," and "there was nothing in particular that was onerous about it."

Support from Institutional Effectiveness Professionals

IE professionals provided training, support, and resources that were identified as important aspects underpinning the APR process. Reviewers felt the IE unit "has the expertise in how to structure [APR]" and the documents and resources IE staff "had prepared in advance were helpful." Participants also shared the materials for assessing the SSR were "all prepared pretty well in terms of laying out specific things that they were looking for in terms of the program that we were going to review" and that IE staff provided "cues as to how to evaluate success."

Emphasis was placed on "having a point person who is really well versed in how [APR] works and isn't afraid to do some of the heavy lifting." In providing a "neutral party administrator-type person who was pulling things together, organizing it, and then sending it back out to us for review" APR was made "relatively painless from the point of view of the internal reviewers." IE professionals assisted by "organizing the meetings, soliciting the original material from the department, facilitating the discussions that we had…helping draft

our report" which freed reviewers up to "be a little bit more thoughtful or contemplative on a higher level" rather than focusing on the minutia of paperwork.

Value in the Form of Learning

In addition to discussing the APR process and support provided, another significant theme that emerged during analysis was how participation in the APR process resulted in learning. All participants acknowledged there was value for the departments under review as they were able, according to one participant, "to get objectively appropriate feedback that you can use to make your department better." The process also resulted in benefits to those serving as IPRs.

"Learning," as a theme, pertained to developing a deeper understanding of APR, other departments within the institution, and how to discuss or engage in assessment within a participant's home department. As one participant shared, "It's always good to have an idea of what these processes are about, otherwise it's like a black box." Another respondent confessed they "came in without very much sense of what it is that we were doing or what we were looking for," but used the opportunity as a learning experience "which is where [they] found the value." There was a desire for many to "see how it worked. To see the criteria the programs are judged by. To see the kinds of materials that were submitted by the program under review." Another participant discussed how their participation gave them a better sense of the purpose of APR: "I saw where the weaknesses were, not necessarily just in their program, but in the way they presented their program. And it helped me to think about what it is that program review should be for." These results support Novodvorsky et al.'s (2015) findings that developing an understanding of the purpose of APR, how it is conducted, and what type of data are utilized are vital to building a positive culture of assessment within an institution.

All participants acknowledged there was value for the departments under review as they were able, according to one participant, "to get objectively appropriate feedback that you can use to make your department better." The process also resulted in benefits to those serving as IPRs.

As suggested by Banta (2014), learning about other departments within the institution was viewed as important to the majority of participants. One participant shared that "being a peer reviewer gave me the opportunity to find out how another department did certain things, like annual reviews, strategic planning, teaching assignments, [and] advising." Another reviewer, who reported being familiar with the department under review, disclosed that "as we worked through the specific list of things that were covered in the review, there was a lot that I didn't know…It ended up being worthwhile for that reason." In addition to giving participants "a decent sense of what unique issues other departments are dealing with" and those issues that many departments share, the review process also "gives you insight into the bigger picture" of an institution.

> I just think the more you learn about a university the better, especially in these days when you know collaboration and cross-disciplinary initiatives and interdisciplinary initiatives are being stressed by the university it is helpful to have a wider knowledge of the university and not to be in your silo, to appreciate and understand what people are doing.

This "opportunity to learn about operations and aspirations of another department" is useful because "you're going back to your home program with kind of a new perspective."

Reviewers also described the manner in which they used their participation in APR as a way to view, discuss, or engage in assessment in their home departments. Ketunnen (2010) asserts that faculty involved in improvement-focused APR will not only produce more meaningful outcomes during the process, but disseminate best practices and contribute to the overall culture of assessment within an institution. One participant stated, "I think when you do it for other departments, it helps you to think about your own program and your own department." Participants mentioned having "a better sense of what to expect," developing "sort of a feel from the other side" of the process, and the ability to "contextualize the way my own program looks at itself." One respondent discussed how through working on the peer review team they were "in a pretty good position to have a template that I can follow of putting together the documents that a review team wants to work with." Serving on

the review team provided some faculty members with an opportunity to discuss different approaches to "a problem [their] own program faces" and then share that with leadership in their home departments.

> The [program being reviewed] was so special because it did have components from all these other different departments. That was kind of interesting to me, and again something that was valuable because of my experience in administering a graduate program that also brings in diverse faculty from many disciplines.

Gentemann et al. (1994) suggested that participation in a well-supported APR may increase the likelihood that departments will seek out the support and expertise of assessment professionals. One participant commented that they "walked away from [the APR] wishing that [their] own department would use [IE professionals, resources, etc.] more effectively… there's a tremendous value there."

Minor Themes

Serving on the review team provided some faculty members with an opportunity to discuss different approaches to "a problem [their] own program faces" and then share that with leadership in their home departments.

Two additional themes that emerged during analysis, albeit to a lesser extent, were cross-disciplinary interactions and service to others. Eight respondents mentioned cross-disciplinary interactions as a valuable part of their experience. When discussing potential benefits of participation in APR, Banta (2014) identified cross-disciplinary interactions as a way to promote professional development. One respondent reported "every time I do anything that's cross-disciplinary, cross-college, cross-university – I get so much out of it." While many people "enjoy meeting colleagues" they don't already know, there is often added value in those interactions. "You learn where the resources are and who is doing what." Another faculty member shared how they "got to know a completely different world" and because of that "reached out to one of the other members of our committee as a potential committee member for one of my students."

While acknowledging their participation in APR would not likely be valued for promotion and tenure, four participants either identified themselves as "service-oriented" or described their involvement in APR as university service. One participant, a full professor, said, "I still try and contribute even though I'm never going to be promoted again." Another participant felt that service did not need to be recognized in order to be valuable; rather the value came from assisting others in your institution:

> To the extent that we can do that with and for each other within the university, I think that's an important thing to do and I'm somebody who believes in putting your money where your mouth is…If I'm going to say that something is important, then I should be involved in it.

Recruitment, Recommendations to Others, and Return Service

Recruitment for service activities in higher education can often be difficult given that some "faculty are very hesitant to get involved in things that don't count for promotion and tenure." Three ideas dominated participant suggestions for increasing recruitment of IPRs: (1) recruitment by senior administrators, (2) personal requests, and (3) clarity about the process. Some faculty felt calling on deans and department heads to nominate potential reviewers for the peer reviewer pool was key "because they should know what their faculty are doing. They should know what the faculty course load is…or if they have a little bit more time that semester than they normally do." Many participants mentioned their appreciation for receiving personal requests from IE staff, as "an invitation is always more appealing, it's easy to ignore something that is generic. It's harder to ignore an invitation."

Clarity about the process was the primary suggestion for improving recruitment of peer reviewers. Participants felt APR was important and that IE staff should "help the whole campus know what you're doing…that the university cares about assessment." Other suggestions include "explaining what [APR] is" and that "it's not painful, it's not punitive. We're trying to figure out a way to help you make your program better and to help emphasize all the great things you're already doing." This coincides with communicating the level of

structured support and continuous involvement of IE staff. In keeping with Ndoye and Parker (2010), APR must aim to highlight student learning and positive programmatic changes in order to develop faculty commitment to assessment and the promotion of a positive culture of assessment.

The majority (93%) of participants indicated they would be willing to serve on a peer review team in the future. However, one faculty member within the 93% did indicate they would only participate again if directly requested by a department under review. Trullen and Rodriquez (2013) and Rodgers et al. (2013) found participation in a meaningful and improvement-driven assessment activity increased the likelihood that those faculty members would acknowledge the value of and engage in other assessment practices in the future. When asked to expand on their willingness to participate as an IPR again, many commented on how much they enjoyed service opportunities that exposed them to different parts of the university. Although a few mentioned the monetary compensation, for many their willingness to participate was rooted in the value they placed on assessment generally and the APR process specifically, as well as a desire to "do something that can make for a positive outcome."

[P]articipants voiced concerns about the extent to which APR reports are reviewed by the respective college dean and the extent to which suggestions provided by the APR internal peer review team are discussed and/or implemented.

Challenges Experienced During the Review Process

Faculty members who served as IPRs offered feedback about challenges they faced. Five participants voiced a concern that the value placed on assessment by the department and college administrators should be more manifest in the process. One of these participants shared, "I think a lot of assessments just die on the vine." These participants voiced concerns about the extent to which APR reports are reviewed by the respective college dean and the extent to which suggestions provided by the APR internal peer review team are discussed and/or implemented.

The majority of participants reported challenges related to the department being reviewed. There were several comments about the development of the SSR. Reviewers wanted to see more transparency in how departments created this document, such as who was involved in its development and what methods of data collection were used to inform its content. One faculty member asserted it is "too easy for the department head to paint a very different picture than as experienced by everyone else in the department." A second participant reiterated that "if problems exist in the department and if you can't uncover them, then you can't fix them." Many commented on the desire to talk to department members beyond the department chair. As one respondent commented, APR is "a community activity. It's not the ideas and visions of a single person, but is part of the larger vision" and there needs to be "community ownership" of not just the SSR, but of the entire process and its outcomes.

Some IPRs felt at loose ends about what happened after their role in the process ended. One participant asked, "What did my time and effort really mean?" These participants wanted to know more about the outcomes of the review or at least receive some type of assurance that the department gained something from the APR process. "There's an emptiness of not knowing…did something wonderful happen because of some suggestions that were made or was it all ignored?" While a department's confidentiality must be kept intact, Linkon (2005) does warn how a perception of there being little return on investment can result in a resistance to assessment activities.

Conclusion

This qualitative study explored the experiences of faculty members who participated as IPRs for academic programs outside of their home department. Study findings suggest the following four elements are important in implementing a meaningful APR process: (1) the process should be well-structured, (2) the process should be improvement-focused, (3) peer review teams should incorporate diverse perspectives, and (4) peer review teams should receive appropriate support from administrative staff. In this study, with IE professionals providing significant support throughout the APR process, IPRs gained multiple benefits.

They (1) learned more about the APR process, (2) learned more about the work of other departments on campus, and (3) learned how to more effectively discuss and engage in assessment within their home departments. The additional themes of cross-disciplinary interactions and service to others were also considered. The results of this study also included insights about participation and recruitment. Challenges discussed by participants will contribute to the ongoing improvement of Virginia Tech's APR process.

There are always considerations when selecting internal peer reviewers vs. external peer reviewers. External reviewers can offer discipline-specific knowledge and insights that may promote innovation, while IPRs hold institutional knowledge that can help inform the feasibility of certain recommendations. The utilization of IPRs has streamlined the APR process at Virginia Tech in comparison to a traditional two- or three-day site visit conducted by a team of external peer reviewers comprised of individuals from multiple institutions. Since utilizing IPRs is considerably less expensive than utilizing external peer reviewers, this could be an effective strategy for APR in an era of declining budgets for higher education. The findings from this study suggest that engaging faculty as IPRs is not only a viable alternative to utilizing external peer reviewers for APR, but one that can be meaningful for the faculty who participate.

This research proposes that engaging faculty as IPRs in the APR process is one strategy for improving faculty perceptions of and participation in assessment activities, which in turn may contribute to a deeper understanding of the review process, influence changes within reviewers' home programs, and increase the dissemination of best practices for assessment. All of these factors contribute to building a positive culture of assessment within an institution of higher education. It is our hope that future research will more closely examine the impact that IPR participation in APR has on perceptions of and attitudes toward assessment practices.

References

Auerbach, C., & Silverstein, L. B. (2003). *Qualitative data: An introduction to coding and analysis*. New York, NY: NYU Press.

Banta, T. (2014). Program review as a catalyst for change. *Assessment Update, 26*(5), 3-4. https://doi.org/10.1002/au.2007

Barak, R. J. (1982). *Program review in higher education: Within and without*. An NCHEMS Executive Overview. Boulder, CO: National Center for Higher Education Management Systems. https://files.eric.ed.gov/fulltext/ED246829.pdf

Bloom, M. F. (2010). Peer review of program assessment efforts: One strategy, multiple gains. *Assessment Update, 22*(5), 5-16. https://doi.org/10.1002/au.225

Blumberg, P. (2017). Academic program assessment: A new expertise for faculty to develop. *The Journal of Faculty Development, 31*(3), 31-38. https://eric.ed.gov/?id=EJ1155897

Bowker, L. (2016). Language and quality assurance: A case study highlighting the effects of power, resistance, and countertactics in academic program reviews. *TTR: Traduction, Terminologie et Rédaction, 29*(2), 177-193. https://doi.org/10.7202/1051018ar

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Thousand Oaks, CA: Sage Publications.

Bresciani, M. J. (2006). *Outcomes-based academic and co-curricular program review*. Sterling, VA: Stylus Publishing, Inc.

Colón, E. P., & Dana, T. (2015). Continuous program improvement: A work in progress. *Assessment Update, 27*(5), 1-16. https://doi.org/10.1002/au.30032

Creamer, D. G., & Janosik, S. M. (1999). Academic program approval and review practices. *Education Policy Analysis Archives, 7*, 23. https://doi.org/10.14507/epaa.v7n23.1999

Creswell, J. W. (2013). *Qualitative inquiry & research design: Choosing among five approaches* (3rd ed.). Thousand Oaks, CA: Sage.

DiBiasio, D. A., & Ecker, G. (1982). *Academic program review in a loosely coupled system*. Paper presented at the Annual Meeting of the American Educational Research Association, New York, NY. https://catalogue.nla.gov.au/Record/5412528

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. https://doi.org/gcph25

Flaherty, R. W. (2016). Abandoning the assessment diet: A lifestyle approach to academic program improvement. *Assessment Update, 28*(1), 7-8. https://doi.org/10.1002/au.30046

Gentemann, K. M., Fletcher, J. J., & Potter, D. L. (1994). Refocusing the academic program review on student learning: The role of assessment. *New Directions for Institutional Research, 1994*(84), 31-46. https://doi/org/10.1002/ir.37019948405

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine.

Jones, S. R., Torres, V., & Arminio, J. (2014). *Negotiating the complexities of qualitative research in higher education: Fundamental elements and issues* (2nd ed.). New York, NY: Routledge.

Ketunnen, J. (2010). Cross-evaluation of degree programmes in higher education. *Quality Assurance in Education: An International Perspective, 18*(1), 34-46. https://doi.org/ddn3s6

Kuh, G. D., Jankowski, N., Ikenberry, S. O., & Kinzie, J. L. (2014). *Knowing what students know and can do: The current state of student learning outcomes assessment in US colleges and universities*. Champaign, IL: National Institute for Learning Outcomes Assessment. https://tinyurl.com/yyeltr39

Linkon, S. L. (2005). How can assessment work for us? *Academe, 91*(4), 28. https://tinyurl.com/y63rvkc5

MacQueen, K. M., McLellan, E., Kay, K., & Milstein, B. (1998). Codebook development for team-based qualitative analysis. *Cam Journal, 10*(2), 31-36. https://tinyurl.com/y23y52xr

Maki, P. (2004). *Assessing for student learning: Building a sustainable commitment across the institution*. Sterling, VA: Stylus Publishing, LLC.

McGowan, V. (2019). Not too small to be strategic: The state of academic program review guidelines and instrumentation in public institutions. *Administrative Issues Journal: Connecting Education, Practice, and Research, 9*(1), 53-67. https://doi.org/10.5929/9.1.1

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. San Francisco, CA: Jossey-Bass.

Ndoye, A., & Parker, M. A. (2010). Creating and sustaining a culture of assessment: The guiding principles of institutions that have established a culture of assessment can be described as internally driven. *Planning for Higher Education, 38*(2), 28. https://tinyurl.com/y7xfn6ps

Novodvorsky, I., Tomanek, D., Foor, R. M., & Burd, G. D. (2015). Supporting and evaluating learning outcomes assessment through academic program reviews. *Assessment Update, 27*(4), 4-16. https://doi.org/10.1002/au.30027

Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Pringle, C., & Michel, M. (2007). Assessment practices in AACSB-accredited business schools. *Journal of Education for Business, 82*(4), 202-211. https://doi.org/10.3200/JOEB.82.4.202-211

Rickards, W. H., & Stitt-Bergh, M. (2016). Higher education evaluation, assessment, and faculty engagement. *New Directions for Evaluation, 2016*(151), 11-20. https://doi.org/10.1002/ev.20200

Rodgers, M., Grays, M., Fulcher, K., & Jurich, D. (2013). Improving academic program assessment: A mixed methods study. *Innovative Higher Education, 38*(5), 383-395. https://doi.org/10.1007/s10755-012-9245-9

Rodrigues, R. (2002). *Want campus buy-in for your assessment efforts? Find out what's important to your faculty members and involve them throughout the process*. Paper presented at the annual meeting of the American Association for Higher Education. http://www.aahea.org/articles/2002-10-feature02_1.htm

Shambaugh, N. (2017). Ongoing and systematic academic program review. In S. Mukerji & P. Tripathi (Eds.), *Handbook of research on administration, policy, and leadership in higher education* (pp. 141-156). IGI Global. https://doi.org/fbxn

Shavelson, R. (2010). *Measuring college learning responsibly: Accountability in a new era*. Stanford, CA: Stanford University Press.

Townley, B., Cooper, D. J., & Oakes, L. S. (2003). Performance measures and the rationalization of organizations. *Organization Studies, 24*(7), 1045-1071. https://doi.org/10.1177/01708406030247003

Trullen, J., & Rodríguez, S. (2013). Faculty perceptions of instrumental and improvement reasons behind quality assessments in higher education: The roles of participation and identification. *Studies in Higher Education, 38*(5), 678-692. https://doi.org/10.1080/03075079.2011.590587

Wergin, J. F. (1999). Evaluating department achievements: Consequences for the work of faculty. *AAHE Bulletin, 52*(4), 3-6. https://www.aahea.org/articles/dec99f1.htm

Wergin, J. F., & Swingen, J. N. (2000). *Departmental assessment: How some campuses are effectively evaluating the collective work of faculty*. Washington, DC: American Association for Higher Education. https://tinyurl.com/ybd4j5qq

Appendix A

APR Study: Interview Guide

1. What previous experience do you have serving as an internal peer reviewer for academic program review or another peer review process? An external peer reviewer (e.g., serving on a discipline-specific accreditation review team)?
2. Please tell me about your experience as a peer reviewer for Virginia Tech's Academic Program Review process.

   What made you decide to participate?

   What was the most valuable aspect of participating?

   Least valuable aspect?
3. What do you feel are the pros and cons of internal peer review vs. external peer review?
4. How might the internal peer review process be improved?

   More meaningful for internal peer reviewers?

   More meaningful for departments being reviewed?

   Ideal composition of internal review teams (e.g., number of reviewers, disciplines, etc.)

   How should internal peer reviewers be matched with departments?

   Suggestions for recruiting faculty members to serve as internal peer reviewers (e.g., incentives)?
5. Would you serve as an internal peer reviewer again? Why or why not?
6. Would you recommend that other faculty members at Virginia Tech serve as internal reviewers? Why or why not?
7. What else do you feel might be important for us to know about the peer review process or your experience as a peer reviewer?

Appendix B

APR Demographic Questionnaire

Participant name:

Preferred pseudonym:

Primary academic department/school:

Primary academic discipline:

Faculty rank:

Number of years you have worked in higher education:

Gender:

Race/ethnicity:

After the interview, the audio recordings will be used to create a detailed transcript of the interview. Once the transcript is complete, you will be contacted and invited to read the transcript and make comments. You will also be invited to review a draft of the research findings.

Preferred email address:

Abstract

In an age of increased focus on improving the student learning experience and engaging in quality improvement processes within academic programs, higher education institutions need to clearly articulate expectations and criteria for the assessment of students' learning to support academic programs in achieving their goals. Given the need for a standardized and transparent process along with an evaluation tool to provide programmatic feedback, the institutional assessment rubric, ASSESS-IT, rubric was developed. This article describes the evidence-based consensus process used to develop and refine ASSESS-IT. ASSESS-IT development process occurred over three phases of revision over a two-year period. The process of rigorous rubric development helped to track the assessment for student learning at the institution, and also led to increased stakeholder engagement, educational development of professionals, and increased meaningful alignment of assessment activities to learning outcomes. The authors provide ASSESS-IT development process and template as an institutional model in which to build and revise other institutional models.

AUTHORS

Constance Tucker, Ph.D.
Oregon Health & Science University

Sarah Drummond, Ed.D.
Oregon Health & Science University

Tanya L. Ostrogorsky, Ed.D.
Oregon State University

CORRESPONDENCE

Email
tuckeco@ohsu.edu

ASSESS-IT: The Development Story of an Institutional Rubric for Evaluating Programmatic Assessment Plans and Reports

In an age of increased focus on improving the student learning experience and engaging in quality improvement processes within academic programs, higher education institutions need to clearly articulate expectations and criteria for the assessment of student learning to support academic programs in achieving their goals (Banta & Palomba, 2014; Council for Higher Education Accreditation, 2019; Montenegro & Jankowski, 2017; Suskie, 2014). Achievement of educational goals requires the implementation of systematic teaching and learning processes to draw inferences and evaluate outcomes, typically referred to as assessment within the academy (Yudkowsky, Park, & Downing, 2019). When done well, the systematic institutional process of assessment of educational outcomes is transparent and allows for an appraisal of outcomes from internal and external audiences (National Institute for Learning Outcomes Assessment, 2011). These audiences include, and center, faculty and students across campus to share their respective student learning outcomes, to review, and to provide programmatic feedback to academic peers. These activities are guided by well-constructed institutional rubrics that can provide a framework to guide the development and evaluation of quality assessment reports (Fulcher & Orem, 2010; Groover et al., 2019; Wicinski, et al 2020). While an abundance of institutional rubrics can be found on higher education websites, there is limited existence of institutional rubrics as well as how they were developed within peer reviewed literature (The State University of New York, 2020; University of Delaware, 2020; University of Hawai'i at Mānoa,

2020). Given the need for a peer-reviewed standardized and transparent process along with a rigorous evaluation tool to provide programmatic feedback, the institutional assessment rubric, ASSESS-IT, was developed.

The ongoing development of an instructional-level rubric for assessment processes requires the strategic engagement of faculty as they are the core drivers of assessment of student learning within their courses and academic programs. Faculty are also responsible for ensuring alignment of programmatic requirements with university core educational competencies (Hutchings, 2010). Many universities have faculty-driven assessment councils that track and advocate for quality assessment institutionally and who also ensure that the institution meets the expectations of external accreditors. For the purposes of this article, the Assessment Council is an inclusive and highly engaged group of faculty, staff, and students who represent all academic programs, student services, institutional research, accreditation, and academic policy. Assessment Council is charged with establishing, monitoring, and reporting on institutional assessment activities for all academic programs (Northwest Commission on Colleges and Universities, 2019, 2020).

To meet that charge, the Assessment Council developed and uses ASSESS-IT to annually evaluate the quality of each academic program's assessment plan and related report to provide feedback for continuous improvement. These reviews include ensuring accurate general program descriptions and contacts, purpose statements, student learning outcomes, and alignment of student learning outcomes with the university's graduation core competencies. This process drives and informs strategic initiatives to improve the assessment of core competencies and student learning outcomes across the institution. This article describes the evidence-based consensus process used to develop and refine ASSESS-IT.

Methods

The ongoing development of ASSESS-IT is evidence of practice-based and process-oriented improvement as the Assessment Council evaluates programs' assessment activities. ASSESS-IT utilizes and reinforces evaluative criteria, definitions of levels, and scoring strategies (Dawson, 2017; Simper, 2018). To develop ASSESS-IT, a systematic literature review was used to build a framework, and the Assessment Council came to a consensus around the minimum amount of evidence needed to make a judgment about quality (Alsina et al., 2017; Dawson, 2017; Moskal & Leydens, 2000; Timmerman et al., 2011; Wald et al., 2012). The literature review included peer-reviewed literature, as well as institution-specific data to incorporate both a top-down and bottom-up approach (Alsina et al., 2017; Dawson, 2017; Goodwin & Leech, 2003; Jonsson & Svingby, 2007; Moskal & Leydens, 2000; Simper, 2018; Timmerman et al., 2011; Wald et al., 2012).

The ongoing development of an instructional-level rubric for assessment processes requires the strategic engagement of faculty as they are the core drivers of assessment of student learning within their courses and academic programs.

The ASSESS-IT development process occurred over three phases of revision over a two-year period. In the first phase, ASSESS-IT was developed from both literature review and institutional data, as well as polling of the Assessment Council members for content items. Throughout the second phase, Assessment Council members offered feedback and negotiated revisions to the content and organization of the rubric through a consensus driven decision making process (National Institute for Learning Outcomes Assessment, 2018). The use of this consensus driven decision making process increased the content validity of ASSESS-IT (i.e., how well the rubric represents the criteria it is intended to evaluate). During this phase of the process, there were numerous opportunities to share opinions via email with the chair and post questions to the group via a consensus driven decision making process, but none were completely anonymous. Finally, in the third phase, council members applied the rubrics to departmental plans and reports and commented on requested edits for ASSESS-IT in both structure and content. The following section will review the process used to develop the current version, as well as, describe the three phases of revision.

Instrument Development

Original Rubric

The rubric's evolution was an attempt to move from an SLO compliance activity to a renewed focus on the use of evidence of student learning.

The original rubric focused on five evaluative criteria for developing student learning outcomes: measurable, student-centered, clear, and aligned to degree and program type. This version allowed programs to focus their attention on the quality of their student learning outcomes. With a strong foundation of quality student learning outcomes, the Assessment Council rubric continued to evolve to include the number of SLOs, alignment with Bloom's taxonomy, and the use of stems to guide action-oriented SLO statements. The rubric's evolution was an attempt to move from an SLO compliance activity to a renewed focus on the use of evidence of student learning.

Iteration One

The first iteration was focused on determining the content and the Assessment Council process for evaluating academic assessment activity (American Educational Research Association American Psychological Association & National Council on Measurement in Education, 2014; Goodwin & Leech, 2003; Timmerman et al., 2011). While external accrediting organizations require continuous assessment processes, it is up to institutions to set individual goals for programmatic assessment. ASSESS-IT was designed to include content items Assessment Council felt were necessary to review the program student learning outcomes. ASSESS-IT's evaluative criteria were re-evaluated to ensure their feasibility and appropriateness across schools and programs. For example, Assessment Council members provided anonymous feedback on the ease of use and quality of information on the rubric as they reviewed academic programs. Techniques to ensure that ASSESS-IT measures what it is intended to measure, i.e. content validity, included the systematic review of the literature related to the evaluative criteria, quality levels, and scoring strategy including guidelines from regional accreditation bodies, the National Institute for Learning Outcomes Assessment, as well as, input from subject matter experts on the university Assessment Council.

Iteration Two

In the second iteration of ASSESS-IT, the Assessment Council intentionally revised the rubric to remedy problems identified within the evaluative criteria descriptions and categories to make it easier to identify distinct quality levels of assessment excellence between programs (Dawson, 2017; Goodwin & Leech, 2003; Jonsson & Svingby, 2007; Moskal & Leydens, 2000). Also, Assessment Council members reflected on the rubric's ability to provide evidence of closing the loop using a consensus driven decision making process (Glassman, et al., 2014; Lennertz & Lutzenhiser, 2006). The results of these sessions were collected and distributed to the Assessment Council for approval through a blind e-vote. Subsequent versions of the rubric were developed to incorporate the refined criterion, and the Assessment Council repeated the process of consensus driven decision making about the language until the final rubric was achieved.

Final Rubric

In the third phase, the rubric went through two additional modifications related to assessment reporting to improve content validity utilizing an informed research approach to stakeholder engagement. First, the closing of the loop dimension was expanded and clarified to include two focused dimensions of using data to inform curricular change: closing the loop using course improvement data or course evaluation feedback and closing the loop using Assessment Council feedback. Second, the Assessment Council added an optional dimension for the submission of a sample rubric so that programs could be recognized for exemplary practices but also so that the Assessment Council could create an assessment repository for educators. These changes reinforced the rubric's validity and demonstrated the value of institutional frameworks to model assessment excellence.

Results

The current ASSESS-IT rubric is divided into two sections: Assessment Planning and Assessment Reporting (Table 1). The assessment-planning rubric includes five dimensions: Communication of SLOs, Progression, Measurable SLOs, Alignment of Core Competencies to SLOs, and Levels of Evaluation Outcomes. The ASSESS-IT reporting section highlights six dimensions, including interpretation of targets, met or not met, engagement of stakeholders in program assessment planning & reporting, closing the loop utilizing course improvement or course evaluation feedback, and Assessment Council feedback, and inclusion of a sample rubric. Table 1 provides the rubric dimensions, definitions when operationalized, as well as a general source citation for each dimension. Discussion and Lessons Learned

This article reviewed the evidence-based consensus process to develop ASSESS-IT. The process of rigorous rubric development helped to track the assessment for student learning at the institution, but also led to increased stakeholder engagement, educational development of professionals, and increased meaningful alignment of assessment activities to learning outcomes.

Stakeholder Engagement: Rubric Development Process

The development of ASSESS-IT is rooted in high levels of institutional stakeholder engagement. Faculty, students, and staff from across the institution provided feedback to the ASSESS-IT development process as members of the Assessment Council and during public forums (i.e., Assessment Academy, Board of Directors Meetings, Student Council, faculty curriculum meetings). At the institutional level, stakeholders focused on the co-creation of meaningful assessment terminology. In addition, faculty, staff, and students engaged, collaborated, and co-created the metrics on the rubric to define engagement at programmatic levels. ASSESS-IT's development supported stakeholder engagement to not only co-create but also evaluate program quality.

Stakeholder Engagement: Adding in a Rubric Dimension

Within academic programs, stakeholder engagement was evaluated as both a quantity and quality metric. Stakeholder quantity defined the variety of groups involved and the frequency of involvement while stakeholder quality identified the participation of the groups in academic programs from information sharing to collaboration. For example, academic programs demonstrated that they engaged employers, students, faculty, staff, and alumni, the frequency of those interactions, and how the information was used to improve student learning. ASSESS-IT development process reinforced and rewarded programs for stakeholder engagement and stimulated discussion on who is and is not engaged in programmatic assessment.

Rubric as a Teaching Tool

While not unique to this rubric, ASSESS-IT serves as a faculty development tool regarding the assessment of student learning and effective curriculum development in two ways. First, faculty from across the institution have clear criteria and standards to evaluate programmatic assessment activities that align with institutional expectations (Andrade, 2000). Second, by summarizing the findings of the ASSESS-IT and providing the aggregated results to faculty, a shared understanding of assessment, direct and indirect methods, levels of assessment, and exemplar tools across programs can be developed. Programs receive feedback on their individual reports with commendations, recommendations, or required changes along with the rubric language to explain the results. Additionally, the Assessment Council provides university best practices and overall results for each dimension so programs can benchmark against the university and gain insight into methods for improvement in assessment and reporting (Oregon Health & Science University, 2020).

The ASSESS-IT development process reinforced and rewarded programs for stakeholder engagement and stimulated discussion on who is and is not engaged in programmatic assessment.

Importance of Meaningful Rubric and Curricular Alignment

The ability of ASSESS-IT to be effective at the institutional level is dependent on the intentional and meaningful alignment of the institutional rubric with instruction and

Table 1
ASSESS-IT Rubric Elements and Standard for Exemplary Rating

Dimension	Exemplary Definition	Reference
Communication of SLOs	Student learning outcomes statements are prominently posted on the institutional website and made available to students.	Excellence in Assessment Rubric (National Institute for Learning Outcomes Assessment, 2019).
Progression	The difference between unique degree/certificate levels is clearly defined in the SLOs, if applicable.	Accreditation standards (Northwest Commission on Colleges and Universities, 2019, 2020).
Measurable SLOs	SLOs are measurable.	S.M.A.R.T. goals (Doran, 1981).
Alignment of Core Competencies to SLOs	Alignment of SLOs with OHSU core competencies is clear	Alignment of Standards And Assessments as an Accountability Criterion (La Marca, 2001).
Levels of Evaluation Outcomes	Assessment methods are appropriately aligned.	Moore's Outcome Framework (Moore, Green, & Gallis, 2009). Blooms' Taxonomy (Bloom et al., 2001)
Interpretation of Targets Met/Not Met	Program explores learner achievement by reviewing and interpreting their targets through a process of data analysis, comparison to peers, and discussion. This includes reflection about missed targets that could prompt a course or program change aimed at improving learning.	Using Evidence of Student Learning to Improve Higher Education (Kuh, et.al., 2015)
Engagement of Stakeholders in Program Assessment Planning & Review	Group and individual engagement regularly include representatives from a) faculty; b) staff; c) students; d) alumni; e) employers; f) external stakeholders/ advisory	Stakeholder Engagement Spectrum (Australian Nursing & Midwifery Accreditation Council, 2017).
Closing the Loop: Course Improvement Data or Course Evaluation Feedback	There is evidence that the program collected, analyzed, and used course level assessment data, not limited to course evaluation data, to inform student learning improvement in at least one course.	Excellence in Assessment Rubric (Banta & Blaich, 2011; National Institute for Learning Outcomes Assessment, 2019). Course improvement through evaluation (Cronbach, 2000). Standards for accreditation (Northwest Commission on Colleges and Universities, 2019, 2020). Key factors influencing student achievement: graduation, retention, completion, classroom environment, and student satisfaction (Elliott & Healy, 2001)
Closing the Loop: Evidence of Program Improvement	Assessment data have been analyzed and used to inform and/or improve the program	Establishing Academic Program Priorities (Shirley & Volkwein, 1978)
Closing the Loop: Address Assessment Council Feedback	The program responded to the assessment council's required feedback from previous assessment cycle, and no further required changes are necessary	Managing Quality in Higher Education: An International Perspective on Institutional Assessment and Change (Pennie, 2001; Brennan & Shah, 2000).
Inclusion of Sample Rubric	The program submitted a sample assessment method (i.e., rubric) which is well aligned with a core competency	Assessing outcomes and improving achievement: Tips and tools for using rubrics (Rhodes, 2010).

program assessment activities. Vertical alignment of programmatic assessment activities with ASSESS-IT supports the ability to use data to inform change at the course, program, or institutional level (Liu, Wrobbel, & Blankson, 2010; Vidic & Weitlauf, 2002). To ensure alignment, the Assessment Council engages in Assessment Academy workshops to educate faculty and staff and commits to maintain a diverse membership who advocates for alignment of assessment activities within their respective programs and curricular councils.

The meaningful process work of rubric development is an effective vehicle for educational quality improvement that reinforces transparency with not just what we assess, but how we assess.

Conclusion

Through meaningful collaboration, the Assessment Council developed ASSESS-IT to evaluate programmatic assessment activities across the institution. The development process required the Assessment Council to make explicit their approach to institutional assessment, which included redefining the purpose of institutional assessment, revising graduation core competencies, and restating programmatic expectations across the institution. The authors provide ASSESS-IT development process and template as an institutional model in which to build and revise other institutional models.

The next steps in ASSESS-IT refinement include examining the validity and reliability of scoring the rubric as well as developing stronger alignment with the co-curricular assessment activities. To this end, the authors will examine ASSESS-IT's inter-rater and intra-rater reliability to determine the extent to which the rubric yields consistent results. Through this ongoing rubric development process, the authors encourage others to focus on both the significant outcomes of rigorous rubric development, but also on the process of fostering a culture of assessment through the engagement of diverse stakeholders, educator development, and alignment of assessment and curricular goals and improvements. The meaningful process work of rubric development is an effective vehicle for educational quality improvement that reinforces transparency with not just what we assess, but how we assess.

References

Alsina, Á., Ayllón, S., Colomer, J., Fernández-Peña, R., Fullana, J., Pallisera, M., Pérez-Burriel, M., Serra, L. (2017). Improving and evaluating reflective narratives: A rubric for higher education students. *Teaching and Teacher Education*, *63*, 148-158. https://doi.org/10.1016/j.tate.2016.12.015

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

Australian Nursing & Midwifery Accreditation Council. (2017). *Stakeholder Engagement Framework*. Retrieved from https://www.anmac.org.au/document/stakeholder-engagement-framework

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

Banta, T. W., & Palomba, C. A. (2014). *Assessment essentials: Planning, implementing, and improving assessment in higher education*. John Wiley & Sons.

Bloom, B. S., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman.

Brennan, J., & Shah, T. (2000). *Managing quality in higher education: An international perspective on institutional assessment and change*. Buckingham, England: Organisation for Economic Co-operation and Development, Society for Research into Higher Education & Open University Press.

Council for Higher Education Accreditation. (2019) *Toolkit for CEOs and CAOs: Making the case for quality*. CHEA/CIQG Publication Series. https://files.eric.ed.gov/fulltext/ED603175.pdf

Cronbach, L. J. (2000). Course improvement through evaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation Models: Viewpoints on Educational and Human Services Evaluation* (pp. 235-247). Springer Netherlands. https://doi.org/10.1007/0-306-47559-6_14.

Dawson, P. (2017). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education, 42*(3), 347-360. https://doi.org/10.1080/02602938.2015.1111294

Doran, G. T. (1981). There's a S.M.A.R.T. way to write management's goals and objectives. *Management Review, 70*(11), 35-36.

Elliott, K. M., & Healy, M. A. (2001). Key factors influencing student satisfaction related to recruitment and retention. *Journal of Marketing for Higher Education, 10*(4), 1-11. https://doi.org/10.1300/J050v10n04_01

Fulcher, K., & Orem, C. (2010). Evolving from quantity to quality: A new yardstick for assessment. *Research & Practice in Assessment. 5*, 13-17. http://www.rpajournal.com/dev/wp-content/uploads/2012/05/A25.pdf

Glassman, A. M., Zell, D., & Duron, S. (2014). *Thinking Strategically in Turbulent Times: An Inside View of Strategy Making: An Inside View of Strategy Making*, Armonk, N.Y.: Routledge. https://doi.org/10.4324/9781315698717

Goodwin, L. D., & Leech, N. L. (2003). The meaning of validity in the new standards for educational and psychological testing: Implications for measurement courses. *Measurement and Evaluation in Counseling and Development, 36*(3), 181-191. https://doi.org/10.1080/07481756.2003.11909741

Groover, C., McBryaer, J.S., Cleveland, R., & Riggs, A.J. (2019). Utilization and perceived utility of instiutitonal administrative and student affairs assessment resources. *Research & Practice in Assessment. 14*, 52-64. https://www.rpajournal.com/dev/wp-content/uploads/2019/07/A4.pdf

Hutchings, P. (2010). Opening doors to faculty involvement in assessment. *NILOA Occasional Paper, 4*. https://learningoutcomesassessment.org/documents/PatHutchings_000.pdf

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*(2), 130-144. https://doi.org/https://doi.org/10.1016/j.edurev.2007.05.002

Kuh, G. D., Ikenberry, S. O., Jankowski, N. A., Cain, T. R., Ewell, P., Hutchings, P., & Kinzie, J. (2015). *Using evidence of student learning to improve higher education*. San Francisco, CA: Jossey-Bass.

La Marca, P. M. (2000) Alignment of standards and assessments as an accountability criterion, Practical Assessment, *Research, and Evaluation, 7*(21). https://doi.org/10.7275/ahcr-wg84

Lennertz, W. R., & Lutzenhiser, A. (2006). *The charrette handbook: the essential guide for accelerated, collaborative community planning*. Chicago: APA, American Planning Association.

Liu, M., Wrobbel, D., & Blankson, I. (2010). Rethinking program assessment through the use of program alignment mapping technique. *Communication Teacher, 24*(4), 238-246. https://doi.org/10.1080/17404622.2010.513002

Montenegro, E., & Jankowski, N. A. (2017). Equity and assessment: Moving towards culturally responsive assessment. *Occasional Paper, 29*. https://learningoutcomesassessment.org/documents/OccasionalPaper29.pdf

Moore, D. E., Green, J. S., & Gallis, H. A. (2009). Achieving desired results and improved outcomes: integrating planning and assessment throughout learning activities. *Journal of Continuing Education in the Health Professions, 29*(1), 1-15. https://doi.org/10.1002/chp.20001

Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. Practical Assessment, *Research & Evaluation, 7*(10), 71-81. https://doi.org/10.7275/q7rm-gg74

National Institute for Learning Outcomes Assessment. (2011). *Transparency Framework*. Retrieved from Urbana, IL: https://www.learningoutcomesassessment.org/ourwork/transparency-framework/

National Institute for Learning Outcomes Assessment (2018, February). *The Assignment Charrette Toolkit*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA). https://www.learningoutcomesassessment.org/ourwork/assignment-charrette/#1549481918909-4e924c6d-3b02

National Institute for Learning Outcomes Assessment. (2019). *Excellence in Assessment (EIA) Designation*. Retrieved from https://www.learningoutcomesassessment.org/eia/

Northwest Commission on Colleges and Universities. (2019). *NWCCU Standards*. Retrieved from https://www.nwccu.org/accreditation/standards-policies/standards/.

Northwest Commission on Colleges and Universities. (2020). *2020 Standards for Accreditation*. Retrieved from https://www.nwccu.org/wp-content/uploads/2019/08/2020-NWCCU-ERs-and-Standards-1.pdf.

Oregon Health & Science University. (2020). *Assessment for Educators: Reports*. Retrieved from https://www.ohsu.edu/education/assessment-educators.

Pennie, I. (2001), "Managing Quality in Higher Education: An International Perspective on Institutional Assessment and Change", *Quality Assurance in Education, 9*(2), 116-117. https://doi.org/10.1108/qae.2001.9.2.116.1

Rhodes, T. L. (2010). *Assessing outcomes and improving achievement: Tips and tools for using rubrics*. Washington, DC: Association of American Colleges and Universities.

Shirley, R. C., & Volkwein, J. F. (1978). Establishing academic program priorities. *The Journal of Higher Education, 49*(5), 472-488. http://dx.doi.org/10.2307/1980511

Simper, N. (2018). Rubric authoring tool supporting cognitive skills assessment across an institution. *Teaching & Learning Inquiry, 6*(1), 10-24. http://dx.doi.org/10.20343/teachlearninqu.6.1.3

Suskie, L. (2014). *Five dimensions of quality: A common sense guide to accreditation and accountability*. San Francisco: Jossey-Bass. https://doi.org/10.1353/rhe.2016.0021

The State University of New York. (2020, September 24). *Assessing Institutional Effectiveness Self-Assessment*. Retrieved from https://system.suny.edu/media/suny/content-assets/documents/academic-affairs/assessment/Institutional-Effectiveness-Rubric-Branded.pdf

Timmerman, B. E. C., Strickland, D. C., Johnson, R. L., & Payne, J. R. (2011). Development of a 'universal'rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment & Evaluation in Higher Education, 36*(5), 509-547. https://doi.org/10.1080/02602930903540991

University of Delaware. (2020, September 24). *UD's General Education Rubrics*. Retreived from https://ctal.udel.edu/resources-2/rubrics/

University of Hawai'i at Manoa. (2020, Septebmer 24). *Rubric for Assessment the Quality of Program-level Student Learning Outcomes & Assessment Plans*. Retrieved from http://manoa.hawaii.edu/assessment/resources/rubrics/rubric-for-assessing-the-quality-of-program-level-student-learning-outcomes-assessment-plans/

Vidic, B., & Weitlauf, H. M. (2002). Horizontal and vertical integration of academic disciplines in the medical school curriculum. *Clinical Anatomy: The Official Journal of the American Association of Clinical Anatomists and the British Association of Clinical Anatomists, 15*(3), 233-235. https://doi.org/10.1002/ca.10019

Wald, H. S., Borkan, J. M., Taylor, J. S., Anthony, D., & Reis, S. P. (2012). Fostering and evaluating reflective capacity in medical education: developing the REFLECT rubric for assessing reflective writing. *Academic Medicine, 87*(1), 1-50. https://doi.org/10.1097/acm.0b013e31823b55fa

Wicinski, M.L., Burr, K., Pinkey, G.L. (2020). Assessing assessment: Using a rubric to improve quality and enhance communication. *Assessment Update, 32*(5), 6-14. https://doi.org/10.1002/au.30227

Yudkowsky, R., Park, Y. S., & Downing, S. M. (2019). *Assessment in health professions education*: New York: Routledge. https://doi.org/10.4324/9781138054394