The assessment cycle: Insights from a systematic literature review on broadening participation in engineering and computer science

Teirra K. Holloman1 | Walter C. Lee1 | Jeremi S. London1 | Chaneé D. Hawkins Ash2 | Bevlee A. Watford1

1Department of Engineering Education, Virginia Tech, Blacksburg, Virginia, USA
2Foresight Strategy Solutions, San Francisco, California, USA

Abstract

Background: In the field of engineering education, assessment and evaluation have been given insufficient attention as they relate to broadening participation. We posit that this lack of attention negatively impacts our ability to develop and implement sustainable solutions at scale.

Purpose: The purpose of this article is to explore the presence of an assessment cycle in the literature related to broadening participation in engineering. The assessment cycle refers to the process by which assessment/evaluation is planned, enacted, and used to improve educational processes.

Scope/Method: The scope of this study was influenced by it being situated in a larger study focused on broadening participation of Black Americans in engineering and computer science. We completed a literature map that illuminated a subset of literature broadly related to assessment/evaluation, and subsequently conducted a systematic literature review of 33 publications reporting on assessment/evaluation efforts. Our analysis of these efforts was grounded in a six-stage assessment cycle.

Results: The results of this study highlight common publishing practices related to the assessment cycle in the context of broadening participation. We find that assessment/evaluation is generally published at the program level, focused on student development or academic success as a proxy for program effectiveness, and concentrated on positive claims.

Conclusion: There is room to significantly improve how assessment/evaluation information is published. By highlighting productive and unproductive publishing practices related to assessment/evaluation, this research has important implications for the use and publishing of assessment/evaluation, particularly as it relates to broadening participation in engineering.

Keywords
accountability, assessment, diversity, program evaluation
1 | INTRODUCTION

As the field of engineering aspires to diversify, organizations continue to implement efforts aimed at bolstering the participation of groups that are historically and systematically marginalized (Lichtenstein et al., 2014). This challenge is commonly referred to as “broadening participation in engineering” (BPE). Here, we argue that our ability to develop and implement sustainable solutions in the context of BPE is intertwined with our ability to assess and evaluate existing interventions, programs, and institutions.

Assessment and evaluation are integral to BPE solutions because practitioners continually compare efforts and exchange best practices, relying on insights from others to determine which practices add the most value and should continue being implemented (George et al., 2019). Moreover, because BPE efforts are often overlooked and undervalued, there is also immense pressure on practitioners to secure external funding and retain internal budgets (i.e., soft funding); such work is often devalued due to its association with what is seen as feminized and immaterial labor (Chen et al., 2019). This circumstance raises the stakes, as it relates to accountability and demonstrating return-on-investment (Rincón & George-Jackson, 2016). As noted by Patton et al. (2019), “when making tough fiscal decisions, key diversity efforts are often cut or minimized” (p. 178). This reality results in BPE practitioners being forced to balance the pressure to minimize costs while maintaining and improving the outcomes about which institutions are concerned (Baber, 2015). In short, in the context of BPE, the pressure to assess and evaluate comes from the need to (1) improve one’s own practices and exchange insights with other practitioners, and (2) justify the need for and use of resources.

The need for improved BPE assessment and evaluation is also echoed by national calls in this area. In fact, “inadequate assessments, metrics, and data tracking” were identified as a key impediment in “Surmounting the Barriers: Ethnic Diversity in Engineering Education” (National Academy of Engineering, 2014, p. 1). Performing self-assessments/evaluations of interventions and programs has also been historically recommended to assist in the recruitment, retention, and employment of people from underrepresented groups in engineering, both at the pre-college and undergraduate levels (Holloman et al., 2018). However, little work has been done to advance the scholarship of assessment and evaluation, particularly in the context of BPE.

First, we describe the limited focus on assessment and evaluation in engineering education research. While publication venues central to the field of engineering education (e.g., Journal of Engineering Education, Journal of Women & Minorities in Science and Engineering, Advances in Engineering Education) address topics related to assessment and evaluation (Cooper et al., 2019; Finelli & Froyd, 2019; e.g., Olds et al., 2005), the conversation primarily focuses on collecting and analyzing data (i.e., issues of measurement). The limitations of this focus become evident when considered through the lens of the assessment cycle. The assessment cycle refers to the process by which assessment (and evaluation) is planned, enacted, and used to improve education processes (Kuh et al., 2015). By better understanding how authors document their use of the assessment cycle in the literature, we aim to highlight productive and unproductive ways in which assessment and evaluation are used. Additionally, we provide corresponding

1.1 | Purpose

The purpose of this paper is to explore the assessment cycle and its presence (or absence) in the literature related to broadening participation in engineering and computer science (CS). The assessment cycle refers to the process by which assessment/evaluation is planned, enacted, and used to improve education processes (Kuh et al., 2015). By better understanding how authors document their use of the assessment cycle in the literature, we aim to highlight productive and unproductive ways in which assessment and evaluation are used. Additionally, we provide corresponding
suggestions for future assessment and evaluation work, which are grounded in literature from fields that are germane to this issue (J. T. Brown, 2017). The audience of this work includes (1) practitioners utilizing assessment and evaluation in their programs, (2) education researchers focused on creating and disseminating assessment/evaluation results, (3) funders who request data to inform funding decisions, and (4) other stakeholders who have an interest in conducting or learning about assessment and evaluation.

1.2 Scope and positionality

Before proceeding any further, it is important to situate this study in the positionality and interests of the authors. Each of the five authors identify as Black/African American, and the inspiration for this literature review comes from a larger study focused on the state of broadening participation of Black Americans in engineering and CS. As part of this larger study, we completed a literature map (J. S. London et al., 2020) that illuminated a subset of literature broadly related to the assessment and evaluation of BPE-related efforts, where Black Americans were the focus or among the participants in the article. This review is scoped to this subset of literature. Due to the exploratory nature of this study, we did not further disaggregate the papers by any attributes of the assessment/evaluation efforts themselves.

Our initial focus on Black Americans was in response to limited progress being made toward reaching racial parity in engineering (Gibbons, 2009; Roy, 2019; Yoder, 2016) despite numerous initiatives focused on Black Americans. For example, we have seen initiatives from the National Society of Black Engineers (NSBE; Fletcher et al., 2017), minority engineering programs at universities across the nation (Lasser & Snelsire, 1996; D. M. Lee & Harmon, 2013; W. C. Lee, 2015), and local chapters of organizations like Black Girls Code and CODE2040 (Bachiri, 2019).

We thought this continued focus was justifiable because BPE practitioners focused on Black Americans often experience the aforementioned pressures to the greatest extent, largely due to Black Americans being the primary target of most attacks on affirmative action. This truth is a byproduct of the US higher education system being originally designed to educate only the White majority and equip White men for leadership roles (Geiger, 2005; Karabel, 2005; Thelin, 2011). In response to desegregation and the civil rights movement, policies such as affirmative action were put in place to aid access to higher education for students from marginalized communities. Unfortunately, many institutions faced and continue to face attacks on affirmative action policies and practices (Rodriguez, 2007; Shehab et al., 2012).

Combined, these two points led us to conclude that Black Americans provided an advantageous lens through which to examine the scholarship of assessment and evaluation in the context of BPE. As a result, many of the decisions we made as researchers were guided by this focus. We acknowledge that our results may have been different had we primarily focused on other populations (e.g., the Latinx community, women, and/or people with disabilities). However, because much of the literature focuses on the population beyond Black Americans, we believe our findings have implications for BPE broadly and discuss the subsequent literature accordingly.

2 DISCUSSING ASSESSMENT, EVALUATION, AND RESEARCH

Because BPE practitioners must regularly communicate with the broader education community, it is important to use language that is commonly accessible. However, finding a common language as it relates to monitoring progress and informing change (or accountability) is difficult because different communities have their own preferred terms. This difficulty is exacerbated in higher education, where there are multiple conflicting approaches to accountability. According to J. T. Brown (2017), there are at least seven fields (e.g., institutional effectiveness, educational measurement) that take a unique approach to this process and many more affiliated associates—such as the Association for the Assessment of Learning in Higher Education (AAHLE), the Association for Higher Education Effectiveness (AHEE), and the American Evaluation Association (AEA). Across these associations, numerous terms are regularly used regarding data collection that informs our decision making, such as assessment, evaluation, institutional research, and institutional effectiveness.

In considering the scholarship of assessment and evaluation, it is important to discuss the complicated relationship among assessment, evaluation, and research. Research is primarily concerned with producing knowledge that is either generalizable or transferable (Tracy, 2010; Winter, 2000). It is the process of inquiry and acquisition of knowledge about a topic or phenomenon. The process is guided by various research methods and methodologies which are ideally informed by clearly identified research questions. The execution of these methods produces data that is analyzed,
interpreted, and used to confirm, deny, or create new knowledge about some theory, topic, or phenomenon. Though this rational model does not always describe the reality of the research process, it does provide a general overview (J. Martin, 1982).

In contrast to research, assessment and evaluation are primarily concerned with gathering useful information to be used for determining the effectiveness of localized efforts (Pellegrino et al., 2014). Though distinct, assessment and evaluation overlap significantly and are often used interchangeably. We acknowledge that the terms assessment and evaluation emerge from different traditions and offer the following definitions to establish a common language for this paper. According to Palomba and Banta (1999), assessment is “the systematic collection, review, and use of information about educational programs undertaken for the purpose of improving student learning and development” (p. 4). In the educational context, it is a mechanism for monitoring progress and informing change, widely perceived as a valuable and essential part of education (Suskie, 2008). According to Allen et al. (2002), evaluation is “the use of assessment findings (evidence/data) to judge program effectiveness; used as a basis for making decisions about program changes or improvement” (p. 1). It aids stakeholders and evaluators in determining if an effort (e.g., initiative, program, course, class) is worthy of adoption, continuation, or expansion (Fitzpatrick et al., 2011).

Given our choice to focus on the collection, review, and use of information about BPE interventions, we hereafter anchor on the process of assessment. This process is iterative and comes in many forms (e.g., quantitative, qualitative), fulfills many purposes (e.g., accountability, improvement), includes many stages (e.g., planning, collecting data, etc.), and is conducted at many levels of analysis (e.g., individual, program, institutions; Leskes, 2002; R. Miller & Leskes, 2005; Town & Pierce, 2013). To discuss the breadth of this process, we incorporate insights related to research and evaluation as well. When discussing the collection of assessment data, we leverage the language commonly associated with research, and when discussing the use of assessment data, we leverage the language commonly associated with evaluation. Further disentanglement of these terms and processes is beyond the scope of this paper.

3 AN OVERVIEW OF ASSESSMENT IN EDUCATION

Before narrowing our focus on BPE, it is important to first understand how assessment is used in education broadly. We will discuss (1) attributes by which assessment can be classified, (2) quality considerations associated with assessment, and (3) barriers to implementation.

3.1 Assessment attributes

For BPE practitioners to accurately assess their efforts, well-thought-out assessment plans are needed. A sound assessment plan has four attributes that should be considered and aligned. The first two attributes relate to the focus of the assessment plan (e.g., what and why?), and include the (1) level and (2) purpose. The next two attributes relate to the implementation of the assessment plan (e.g., how?), and include the (3) function and (4) form. Each attribute is discussed further below.

3.1.1 The focus of assessment plans

There are several approaches to discussing the focus of an assessment plan, either focusing on the different purposes (i.e., why an assessment plan is needed) or the different levels (i.e., aspects of the experience or environment). Regarding the levels of assessment, according to R. Miller and Leskes (2005), there are five different categories: (1) assessing individual student learning within courses, (2) assessing individual student learning across courses, (3) assessing courses, (4) assessing programs, and (5) assessing the institution. Regarding the purpose of assessment, common categories include accountability, improvement (Leskes, 2002), grading, learning, and coaching (Hansen, 2012). It is important to understand the different classifications because they have implications for how assessment questions are formulated. For example, at the individual level, one might ask “How well has the student achieved the learning outcomes set for the course?” whereas at the program level one might ask, “Does the program advance institution-wide goals as planned?” To assist readers with exploring the various levels and purposes of assessment across the literature, we provide an overview in Table 1.
### Table 1  Levels and purposes of assessment

| Classification | Category     | Definition                                                                                                                                                                                                 |
|----------------|--------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Level          | Individuals  | Investigating how well individual participants are learning and meeting predetermined goals within and across courses (R. Miller & Leskes, 2005)                                                                    |
|                | Courses      | Investigating how well a group of participants are learning and meeting predetermined goals within and across courses (R. Miller & Leskes, 2005)                                                         |
|                | Programs     | Investigating how well a program is meeting predetermined goals and outcomes (R. Miller & Leskes, 2005)                                                                                                    |
|                | Institutions | Investigating how well an institution is meeting predetermined goals and outcomes including student learning and stakeholder satisfaction (R. Miller & Leskes, 2005)   |

| Purpose        | Assist learning | Provide specific information about students' strengths and difficulties with learning (National Research Council, 2001)                                                                                      |
|                | Individual achievement | Used to make decisions about individuals and is conducted to help determine whether a student has attained a certain level of competency after completing a particular phase of education (National Research Council, 2001) |
|                | Evaluate program | Help inform decisions about the quality and effectiveness of educational programs and institutions (National Research Council, 2001)                                                                        |
|                | Aptitude       | Used to predict the performance of a student in some future situation (National Research Council, 2001)                                                                                                    |
|                | Accountability | Assessment of some unit to satisfy stakeholders external to the unit itself (Leskes, 2002)                                                                                                               |
|                | Improvement    | Feeds directly, and often immediately, back into revising the course, program or institution to improve student learning results (Leskes, 2002)                                                           |
|                | Grading        | Behaviorism measured student progress via the acquisition of isolated facts (Hansen, 2012)                                                                                                               |
|                | Learning       | Investigates where students are in their process of uncovering knowledge and making it their own (Hansen, 2012)                                                                                             |
|                | Coaching       | When students welcome and expect ongoing feedback to help them make progress (Hansen, 2012)                                                                                                               |

### Table 2  Assessment functions to consider when designing an assessment plan

| Category       | Definition                                                                                     | Timing of data collection |
|----------------|-----------------------------------------------------------------------------------------------|----------------------------|
| Formative      | Gathering information about student learning during the progression of a course or program, usually repeatedly | During                     |
| Summative      | Gathering information at the conclusion of a course, program, or undergraduate career          | End                        |
| Readiness      | Exploring the satisfaction of prerequisite entry skills to remedy entry deficiencies and/or categorize learning groups | Before                     |
| Placement      | Determine necessary course or unit objectives                                                  | Before                     |
| Diagnostic     | Identify common breakdowns in mental models                                                    | During                     |

### 3.1.2  The implementation of assessment plans

Once the level and purpose have been identified, practitioners can determine the function (i.e., how the results will be used) and form (e.g., appropriate methods). To assist readers with exploring the functions and forms to consider, we provide an overview in Tables 2 and 3.

Regarding function, the most common approach to differentiating is to distinguish between formative assessment and summative assessment: formative assessment is focused on making improvements for current students or participants, whereas summative assessment is focused on following cohorts (Leskes, 2002). However, other categories include readiness, placement, and diagnostic. These distinctions are important because they enable educators to identify the focus of measurement, nature of the sample, item difficulty, time of administration, and how the results will be used.
For example, according to M. D. Miller et al. (2008), assessments that occur prior to instruction or intervention can either assist with remedying entry deficiencies, assigning students to groups, informing instructional planning, or determining advanced placement (readiness and placement); an assessment that occurs during instruction can either improve and direct learning through ongoing feedback, or remedy errors related to persistent learning difficulties (formative and diagnostic); and assessment that occurs at the end of instruction can be used to assign grades, certify accomplishment, or evaluate teaching (summative).

The forms of assessment tend to come in five main categories: embedded, direct, indirect, qualitative, and quantitative. Several factors implicate which form of assessment may be most appropriate to use. For example, the assessment form tends to indicate the type of data that will be collected in light of the assessment question. It is important for the assessment form to be aligned with the overarching assessment purpose, questions, and/or objectives. Resources, time constraints, accessibility of data, and analysis tools must also be considered. Lastly, because assessment results are often used to convince stakeholders, it is also important to consider the epistemological frameworks of those who will be receiving the results.

### 3.2 Assessment quality

As previously mentioned, there are some similarities between research and assessment, namely when it comes to quality. We define “quality” as the extent a process exemplifies standards and criteria noted in germane literature. For example, Moskal et al. (2002) note that the only way for engineering educators—and in our case BPE practitioners—to adequately monitor progress is to be aware of the important issues that surround assessment measures. Much like with engineering education research, two major issues related to assessment quality are validity and reliability. Validity refers to the degree to which a measurement measures what it is intended to measure or the degree to which interpretations of results are used correctly. Reliability refers to the degree of consistency of assessment scores (Moskal et al., 2002). These terms are often used when researchers employ quantitative methods. When researchers use qualitative methods, they are likely to use terms like transferability, trustworthiness, and dependability as a criterion by which to judge the work (Golafshani, 2003). Transferability is another way of looking at external validity, trustworthiness at internal validity, and dependability at reliability (Leydens et al., 2004). Ultimately, regardless of the approach, it is imperative that criteria are established, systematically used, and openly discussed in reports and publications.

Considering the pragmatic nature of assessment and pressure for BPE practitioners to demonstrate their value to a variety of external stakeholders, it can be difficult to determine what is considered “good assessment” given the many levels, purposes, forms, and functions. Fortunately, criteria regarding assessment best practices have been well documented over the years. For example, Suskie (2008) provides a concise list of five best practices, which summarizes over a dozen of these various sources. These characteristics state that good assessments (1) are used, (2) are cost effective, (3) yield reasonably accurate and truthful results, (4) are valued, and (5) focus on and flow from clear and important goals. Some suggestions to achieve these characteristics are to consider the purpose of assessment, have collaboration occurring between the student and the teacher, and use the feedback from the assessment when making future decisions. While these suggestions are informative, practitioners must interpret them as needed to fit their varying and ever-changing contexts.

In short, there is no “one-size-fits-all” approach to assessment, and it should not be perceived as such. “To say that assessment is [a] good assessment is akin to saying that a medical test is a good test; each can provide useful

| Category   | Definition                                                                                                                                 |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| Embedded   | Gathering data about student learning in a way that is ongoing and, thus, built into the teaching-learning process                         |
| Direct     | Gathering data that demonstrates the learning objectives/outcomes of interest                                                            |
| Indirect   | Gathering data that explores participant perception about learning and the surrounding environment                                        |
| Qualitative| Collecting and analyzing textual data (surveys, interviews, focus groups, conversational analysis, observation, ethnographies) with emphasis on the context within which the study occurs |
| Quantitative| Collecting and analyzing data often from surveys, with the intent to generalize and make inferences                                      |
information only under certain circumstances” (National Research Council, 2001, p. 222). In other words, assessment can be considered good only when all parts are in alignment, it is implemented with a goal or need in mind, and has been validated and deemed reliable.

### 3.3 The barrier to assessment

Nonetheless, numerous factors make achieving “good assessment” difficult. According to Hoey and Nault (2008), these barriers primarily relate to aspects of trust (e.g., trust in the motives, questions, methods, data, etc.), which they viewed as fundamental to assessment design and implementation. Trust will either facilitate or pose obstacles to change implementation. Accordingly, it is important to foster trust via three main avenues: (1) culture, (2) structure, and (3) leadership. Each of these avenues must operate in a way that promotes assessment to establish stakeholders’ buy-in. For example, an organization’s culture would ideally promote open communication, value evidence-based decision making, offer incentives for improvement, and provide resources for those charged with an assessment. To promote the value of assessment, leaders must also make it a normalized practice within the organization. This process should include using those results in ongoing discussions and/or in creating new initiatives; and allocating funds, training, and/or personnel to aid in conducting assessment and the implementation of results. Unfortunately, for many BPE practitioners, trust is not there. With the overwhelming and constant pressure to show value, it is difficult to report any shortcomings or failures. Additionally, being often among the first to lose internal funding makes it challenging for BPE practitioners to trust that their work is valued.

The scholarship of assessment requires an understanding of how assessment can be operationalized, classified, evaluated, and promoted. Next, it is important to understand how each of these components fit together.

### 4 CONCEPTUAL FRAMEWORK: AN ASSESSMENT CYCLE

To systematically examine the use of assessment, we situate our investigation in a six-stage assessment cycle (Kuh et al., 2015). This assessment cycle is a continual process that begins with developing a plan and identifying desired outcomes, followed by collecting and analyzing data. Next, the assessment results are shared so that necessary change can be identified, implemented, and assessed. An overview of each stage will be discussed in the following sections.

#### 4.1 Step 1: Plan and identify outcomes

In the initial stage of the assessment process, a plan must be established. During this stage, assessment questions and/or objectives should be formulated. Assessment questions set the foundation of the assessment plan, informing the evaluator of the focus of the investigation and what data is to be collected. For every assessment question, Fitzpatrick and colleagues suggest that one or more sources of data be identified. While original data may be necessary, the use of existing data may also be beneficial. However, the existence of data does not automatically mean it must be used. The types of questions asked vary depending on the level of assessment (R. Miller & Leskes, 2005) and should be developed with stakeholders in mind or present (Fitzpatrick et al., 2011). Finally, questions used during an assessment should be important to most, if not, all stakeholders.

#### 4.2 Step 2: Collect data

Once data sources have been identified, the collection of data can begin. According to Worthen, Borg, and White (as cited in Fitzpatrick et al., 2011, p. 348), there are five categories by which data collection can be obtained: (1) directly from individuals identified as sources of information, (2) by an independent observer, (3) by a technological device, (4) with unobtrusive measures, and (5) from existing organizational information or formal repositories or databases. Information collection arrangements must also be made, identifying who will collect data, under what conditions, and when will collection take place. Important deadlines, timelines, and resources should also be considered. Given the focus on collecting data, this step in the cycle often overlaps with education research.
4.3 | Step 3: Analyze data

For each assessment question, evaluators should ensure that they have a plan for analyzing the corresponding data or information. Fitzpatrick et al. (2011) note that it is important to specify how the information will be analyzed. They suggest doing so through two steps. First, one should determine what techniques will be used during analysis (e.g., descriptive statistics, emergent coding). Next, one should then determine by what means these analyses will take place (e.g., Excel, NVivo). When interpreting these results, Fitzpatrick et al. recommend involving stakeholders and clients because varying perspectives can result in varying interpretations. This step also overlaps with education research.

4.4 | Step 4: Share results

When sharing results, it is important to be aware of when this information is needed, who needs it, and in what format. Fitzpatrick et al. (2011) offer a matrix to be considered for each assessment question to help organize this information. It includes understanding the audience, the necessary content, reporting format, date of the report, and the context it will be presented; this level of detail and organization also emphasizes the notion that questions that are to be investigated should be intentional, important, and useful to stakeholders. The timing, audience, and format of the results when they are shared is another notable difference between assessment and research.

4.5 | Step 5: Identify and implement changes

After results have been disseminated, the hardest part of the cycle begins, requiring stakeholders to act on the results. Kuh et al. (2015) attribute this difficulty to an overemphasis on data collection and research design and suggests that emphasis should shift to implementation and assessing impact. Ensuring the utility at the beginning of an investigation will help in mitigating future oversight. Additionally, Kuh and colleagues suggest linking assessment to both internal and external processes such as accreditation and promotion. This step in the cycle also makes it difficult to separate assessment from evaluation because both processes require that the information previously collected be used to advance some desired end.

4.6 | Step 6: Assess the impact of changes

This final stage of the cycle officially closes the assessment loop by ascertaining impact, and by extension, propelling the process into another iteration. Implementing changes does no good if stakeholders are unaware of the positive or negative consequences that resulted.

5 | RESEARCH DESIGN

To explore the use of assessment/evaluation in the context of BPE, we conducted a systematic literature review (SLR). SLRs have been used to summarize, critically evaluate, and reconcile conflicting evidence to inform policy and practice (Borrego et al., 2014). Despite the well-established use of literature reviews in engineering education, SLRs are not used as often (Borrego et al., 2014). The value in an SLR can be found in the varying degrees of its purpose from describing the current state of knowledge to evaluating a theory, to identify gaps in the literature (Petticrew & Roberts, 2006).

In order to conduct this SLR, two primary sources were used to guide our research design. Petticrew and Roberts (2006) explain the details of the process as it relates to various disciplines conducting SLRs, and Borrego et al. (2014) operationalize the SLR method for the engineering education community. The five major steps of this SLR include the following:

1. Formulate guiding research questions and corresponding inclusion criteria.
2. Find and catalog sources.
3. Critique and appraise the quality of selected literature.
4. Address bias, validity, and reliability concerns.
5. Synthesize insights.

These steps will serve as a framework for the structure of the remainder of this section. The following sections will break down the process of collecting and identifying articles as a part of a larger study and identifying those appropriate for this study.

5.1 | Formulating guiding research questions and inclusion criteria

Our exploration of the literature was guided by the following research question: *How are assessment and evaluation practices related to broadening participation communicated in the literature?* The initial inclusion criteria used during the collection of literature are as follows: (1) publication date range, (2) database selection, and (3) search string selection. This paper is part of a larger project, and more about these processes can be found in a previously published article about the overall process of conducting an SLR (Phillips et al., 2017). The date range (e.g., 1970–2017) was informed by historical events associated with efforts to broaden the participation of Black Americans in engineering and CS, specifically the establishment of several national-level efforts focused on BPE in the 1970s. For example, the National Action Council for Minorities in Engineering, Inc. (NACME) was established in 1974 and the NSBE was established in 1975. With the help of a librarian, databases and search strings were identified, as illustrated in Table 4. They helped to limit publication bias by keeping the database options as wide-ranging as possible. Additionally, their expertise was used to identify search strings, alongside input from researchers in our professional network and keywords associated with “Diversity” from the *Engineering Education Research Taxonomy* (Finelli et al., 2015).

5.2 | Finding and cataloging sources

After removing 100 duplicate publications, 1080 articles were evaluated against three additional hierarchical inclusion criteria: (1) Is the article written in English and about education or the science, technology, engineering, and mathematics (STEM) workforce in the United States? (2) Is the article focused on engineering or CS in any context, or STEM

| Database name | Search string | Note |
|---------------|---------------|------|
| Education Source and PsycINFO (EBSCOhost interface) | (bias OR discrimination OR multicultural* OR inclusive* OR racism OR prejudice) OR (motivation OR attainment OR achievement OR aspiration OR persist* OR retention)) AND ((AB african w2 american OR SU african w2 american OR TI african w2 american) OR (AB black OR SU black OR TI black) OR (AB people N2 color* OR SU people N2 color* OR TI people N2 color*)) AND ((AB STEM OR SU STEM OR TI STEM) OR (AB engineer* OR SU engineer* OR TI engineer) OR (AB “computer science” OR SU “computer science” OR TI “computer science”)) | Search all fields for words used to include or exclude people Search abstract, title, subject headings for terms used for African American Search abstract, title, subject headings for STEM, engineering, and computer science |
| Compendex and INSPEC (Ei Village interface) | < ((motivation OR attainment OR achievement OR aspiration OR persist* OR retention) WN All fields) > OR < ((bias or discrimination or multicultural* or inclusive* or racism or prejudice) WN All fields) > AND < (((african ONEAR/2 american) WN KY) OR ((black) WN KY)) OR ((people NEAR/2 color) WN KY) > AND < ((STEM OR engineer* OR “computer science”) WN KY) > | Quick search, autostemming off, search all fields Search subject/title/abstract, autostemming off Search subject/title/abstract, autostemming off |

Abbreviation: STEM, science, technology, engineering, and mathematics.
disciplines in a K–12 context? (3) Is the article focused on issues or the experiences of Black Americans, or on some aspect of the wide variety of topics associated with broadening participation? Criterion 3 allowed for the inclusion of publications that might not focus solely on Black Americans. This inclusion was done to account for the possibility of learning from publications that emphasized Black Americans without solely focusing on them. We intended to be mindful of the realities of targeting a specific group in the current legal landscape and the frequency with which multiple marginalized groups are included in a single intervention. In total, 470 out of 1180 (40%) articles met the criteria and spanned across the education-to-workforce pathway, as illustrated in Figure 1. Additional details about the information in Table 4 is included in a systematic mapping review that has been published elsewhere (J. S. London et al., 2020). This paper focuses on a subset of 51 articles related to assessment.

5.2.1 | Limitations of finding sources

Given that all initiatives, programs, and institutions are not mandated to publish assessment/evaluation results, this review can only speak of the work made available at the time of the literature search. Additionally, we must acknowledge that assessment was not a search term used in data collection, as this study is the result of a larger project that dictated data collection. Lastly, due to our initial interest and the exploratory nature of this study, we did not disaggregate literature based on assessment level or intervention type and we may have overlooked articles focused solely on the Latinx community, women, and/or people with disabilities. Each of these aspects bring limitations to this work and therefore cannot speak to the total landscape of BPE assessment.

5.3 | Critique and appraise quality of literature

Following the documentation of the literature map results, the initial 51 assessment documents underwent a quality check. Four questions were used to appraise all documents being utilized in this SLR. The questions were as follows:
1. Is the problem/purpose/aim of the study clearly stated?
2. Is the sampling strategy apparent and appropriate? Does the sample represent the target population?
3. Is information about data collection procedures apparent and appropriate?
4. Is information about the approach to analyzing data apparent and appropriate for addressing the study's purpose?

Following this final step of appraisal, only 34 documents remained. Unfortunately, one document was not accessible given available resources and thus was discarded from our list of articles to review. This resulted in a final total of 33 assessment documents.

5.4 Address bias, validity, and reliability concerns

To address bias, validity, and reliability while conducting this SLR, our research team consulted works outlining various steps to mitigate these outcomes (Mullen & Ramírez, 2006). To minimize author selection bias, search techniques that consider gray literature (and not just journal publications, for example) were employed (Higgins & Green, 2008) and primary studies were masked for author names, affiliations, and journal names during the quality appraisal. Additionally, a librarian was brought into the process of identifying keywords and databases to mitigate the impact of researcher bias. Furthermore, to help establish the reliability and further minimize our own bias, the criteria for inclusion/exclusion and quality assessment were discussed and applied by multiple coders. Collectively, these efforts were employed to address concerns related to bias, validity, and reliability.

Although efforts were made to minimize bias, limitations remain. According to Cook et al. (1997) the limitations of an SLR stem from (1) the quality and quantity of the selected studies, and (2) the quality of the systematic review procedures. One fault in this study lies in the fact that the literature in this review was collected in January of 2017; therefore publications after that time have been excluded from this review. Lastly, this study is a result of seeking BPE assessment literature across the education pathway and therefore we must acknowledge the varying histories that exist between K–12 and postsecondary education and their use of assessment.

5.5 Synthesize insights

The synthesis step included extracting data and coming up with themes. After the final quality check, data were extracted from each of the 33 documents. For each document, we identified Where (what is the context of the program?), Who (which populations are being used to generate information?), What (what assessment questions or objectives are communicated?), How (what metrics and outcomes are being assessed?), When (what is the temporal, duration, proximity to the experience, the longitudinal nature of the assessment?), Findings (what sort of evaluative claims are people making?), and Implications (how will the results be used?), followed by a critique. A summary table was composed to provide a comprehensive view and summary of each study. These documents were compared across each section of every document and the pros, cons, similarities, and suggestions were identified for improvement based on our conceptual framework of the assessment cycle.

6 RESULTS AND DISCUSSION

Following the data extraction from the final 33 documents, we utilized Kuh and colleagues' assessment cycle to inform our analysis. Five major themes emerged as it relates to the initial stages of the assessment cycle, covering topics related to levels of assessment, purpose, and design. More specifically, Theme 1 focused on where the published assessment tends to occur. Theme 2 describes the sparse use of assessment at some levels. Theme 3 recounts the research designs used in BPE assessment plans. Theme 4 depicts the types of data being collected as a proxy for program performance. Lastly, the kinds of claims that are made based on the results of these assessment plans are discussed in Theme 5. Each theme is discussed below.
6.1 Theme 1: The assessment articles in our study most frequently focused on programs at the K–12 level and predominantly White institutions of higher education, and incorporated students with various marginalized identities

The papers included in this SLR focused on assessments most frequently used in the context of programs at the K–12 level and predominantly White institutions (PWIs) of higher education. Of the 33 articles identified, 9 were at the K–12 level, 20 were at the undergraduate level, 1 was at the graduate level, and 3 included participants across levels. It is important to note that none of the articles focused on community college and workforce segments.

At the K–12 level, some investigations focused on the summer programs available for grade school students (Ladeji-Osias et al., 2016; Oliveira, 2010; Wilson et al., 2015), whereas others looked into programs that coupled a summer experience with continuous resources throughout the following academic year (Clark et al., 2015; Di Salvo et al., 2013; Dunn & Veltman, 1989). Interventions often included aspects of mentoring, academic support such as tutoring, or professional development (Hanifin & Schumack, 2001; Ladeji-Osias et al., 2016; Lam et al., 1997). Other aspects included living-learning communities, seminars, hands-on activities, and periodic group meetings (E. Brown, 2007; Kendricks & Arment, 2011; Ogilvie, 2007). Some programs also featured monetary aspects in the form of scholarships, incentives, or pay for employment (Di Salvo et al., 2013; Ericson et al., 2014).

Of the articles that listed participation requirements for high school seniors and undergraduate participants, most required students to have a particular academic standing to receive or maintain program support (Anderson-Rowland & Ruben, 2008; Hrabowski & Maton, 1995; Kendricks & Arment, 2011; Maton et al., 2000). Contrarily, middle school and high school programs allowed students to participate as long as they had an interest (Di Salvo et al., 2013; Ferreira, 2002) or qualified as low socioeconomic status (SES; Dunn & Veltman, 1989).

Most programs beyond K–12 took place at PWIs, though there were notable exceptions. For example, Kendricks and Arment (2011) reported on an undergraduate program at a Historically Black College and University (HBCU) that adopted a K–12 classroom family model. Another HBCU study was conducted by B. Martin (2014), where he sought to understand if the type of academic advising a student received impacted their overall academic performance and satisfaction with advising.

The papers included in this SLR focused on assessment frequently used by BPE stakeholders that simultaneously focused on various marginalized dimensions of identity. While Black Americans were typically the majority population in the articles, some studies were framed around all racial/ethnic minorities (i.e., Black Americans, Latino/a/x, and Native Americans), women (Donnelly, 2007; Lam et al., 1997; Murphy et al., 2010; Wilson et al., 2015), first-generation or low SES (Dunn & Veltman, 1989; Gardner-McCune et al., 2013; Moyer, 2013). In studies that included the use of White and/or Asian participants, the authors gave explicit explanations as to why this was the case, either due to legal concerns, other requirements that the students met like low SES, or for comparison purposes (Dunn & Veltman, 1989; Litzler et al., 2010; Ogilvie, 2007).

6.2 Theme 2: The assessment articles in our study were most frequently program-level assessment and seldom focused on institution-level assessment or the broader context surrounding underrepresented students

Levels of assessment and their purpose play a major role in guiding the types of questions and topics explored in the assessment. Our analysis found that most articles investigated program effectiveness based on either a direct or indirect measure of student development or performance. Occasionally, some investigations would stray away from the status quo by moving beyond the program level. For example, Litzler et al. (2010) investigated climates in engineering at various universities in the United States using a 132-item survey instrument, while McCullough et al. (2014) created a rubric for students to complete weekly that informed their mentoring sessions. An included dissertation (Jia, 2015) assessed how stricter high school math prep impacted the performance of underrepresented students (i.e., Black Americans, Latinos/as/xs, and women) as compared to White men. While these topics were not focused on Black Americans, results shed light on findings directly related to Black Americans’ persistence in engineering. Lastly, White et al. (2008) presented challenges and issues that arose during the creation of a program evaluation plan, but they did not present the evaluation plan itself.

Moreover, the seemingly narrow focus of assessment also limits the degree to which we understand the broader context within which BPE stakeholders work. If we understand that BPE is a multifaceted complex task, we need to consider our overall understanding of the problem space. When thinking about the context in which Black Americans are situated, there is much to be understood about how the engineering ecosystem does and does not adequately
support them. One means of gathering more information is through assessment at various levels within various contexts. Course-level and institution-level assessment from the perspective of Black American students is necessary. We need to understand the parts of the environment that were not designed for Black American students to succeed just as much as we need to understand the interventions that were created with them in mind. Our understanding of one helps to inform the changes implemented to improve the other.

### 6.3 | **Theme 3: The assessment articles in our study utilized both quasi-experimental designs and nonexperimental designs**

Assessment designs varied across articles and could be categorized into one of five types. Three quasi-experimental designs were frequently employed: (1) pretest–posttest design, (2) posttest-only design, and (3) posttest only design with repeated measures. Additionally, nonexperimental designs emphasized both formative and summative functions of assessment. Table 5 provides an overview of each design including references to a few example articles.

#### 6.3.1 | Posttest only

This design describes assessment plans that involved collecting data at the end of the intervention—sometimes including more than one data type. For example, Donnelly (2007) administered multiple assessment tools after the intervention. Post intervention, data were collected (via surveys, interviews, and focus groups) from the participants to assess the effectiveness of the program. Additionally, faculty who participated in the program completed surveys, providing insights on what they perceive to be important elements of the program.

#### 6.3.2 | Pretest–posttest

This design describes assessment plans that involved collecting data before and after an intervention, typically the same types of data, to demonstrate growth and development. For example, D. Knight et al. (2013) administered pre- and post-surveys to measure identity development in program participants. One survey was administered before and after two summer bridge programs. Additionally, another survey was conducted before and after two first-semester courses. Data were used to compare pre- to post-results, across demographic groups, and across interventions.

#### 6.3.3 | Posttest only design with repeated measures—Cohort overtime

This design describes assessment plans that involved collecting data from a particular cohort of students at the end of an intervention and longitudinally thereafter. At the undergraduate level, for example, some studies would follow

---

**Table 5** Quasi-experimental and nonexperimental designs used in some of the assessment documents

| Design                                      | Description                                                                 | Example articles                                                                 |
|---------------------------------------------|-----------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| Posttest only                               | A single observation of any kind performed after a treatment                 | Donnelly (2007); Gardner-McCune et al. (2013); J. Good et al. (2002); Lee and Cross (2013) |
| Pretest–posttest                            | Two observations of any kind performed before and after a treatment          | Hrabowski and Maton (1995); Clark et al. (2015); D. Knight et al. (2013)         |
| Posttest only with repeated measures        | Performing multiple observations on the same sample over time (i.e., cohort performance over time) | Murphy et al. (2010); Jiang et al. (2005)                                        |
| Nonexperimental—formative                   | Multiple observations of participants during the progression of the intervention | E. Brown (2007); J. M. Good et al. (2000); McCullough et al. (2014); Ogilvie (2007); Wilson et al. (2015) |
| Nonexperimental—summative                   | Multiple observations of an intervention occurring at the end of each intervention cycle | Lam et al. (1997); Kendricks and Arment (2011); Lasser and Snelsire (1996) |
students from their participation in a bridge program to their completion of the first/second year of an engineering program or through graduation. Murphy et al. (2010) observed the graduation rates of program participants and determined the likelihood that someone would graduate if they participated as opposed to those that did not; Jiang et al. (2005) tracked five North Carolina Louis Stokes Alliance for Minority Participation cohorts over a 5-year period.

6.3.4 | Nonexperimental—Formative

This design describes assessment plans that involved observing participants and collecting data throughout the intervention. For example, J. M. Good et al. (2000) analyzed student journal entries that were collected throughout the entire academic year. It was noted that the program coordinator engaged with students and their journal entries weekly to “encourage discourse regarding program development and improvement with each mentor” (J. M. Good et al., 2000, p. 377). Similarly, E. Brown (2007) used measures like student accolades and changes in academic major during the intervention as factors of program success.

6.3.5 | Nonexperimental—Summative

We previously described assessment plans that involved collecting data about a cohort of students at the end of an intervention cycle, repeatedly over time (i.e., posttest only design with repeated measures). The nonexperimental (summative) design is somewhat similar in that data is collected at the end of an intervention. However, the difference for this design is that data are not collected from the same group of participants. Lam et al. (1997) illustrate this. They aimed to better understand a multifaceted program, Increasing Diversity in Engineering Academics (IDEAS), that provides support for students academically, socially, emotionally, finanancially, and professionally year-round throughout their undergraduate careers. Therefore, an analysis of the pros, cons, and problems associated with the IDEAS program was conducted at the end of the intervention cycle. Similarly, Lasser and Snelsire (1996) evaluated the benefits of proactive mentoring over several years, using various measures to represent snapshots of the program over time.

6.4 | Theme 4: The assessment articles in our study most frequently investigated program outcomes using indirect and direct measures of student development or academic success as a proxy for program performance

While a variety of assessment questions were expressed in the articles, most articles sought to determine the impact of an intervention program on participants. This entailed investigating how successful a program was based on the participants’ likelihood of pursuing or persisting in STEM (Anderson-Rowland & Ruben, 2008; Dunn & Veltman, 1989; Wilson et al., 2015), affecting participants’ attitudes and interest toward STEM (Clark et al., 2015), or participants’ self-efficacy in STEM-related areas (Di Salvo et al., 2013).

Programs targeting undergraduate students primarily relied on student performance in subsequent semesters or years (i.e., direct measures) to determine effectiveness. For example, Ogilvie (2007) compared the grade point averages (GPAs) of students in a program against non-participants’ GPAs, using retention rates and student achievements as data sources. Taylor et al. (2008) investigated whether students who participated in a STEM learning community earned higher grades, had improved conceptual understanding of STEM material, or had higher retention rates in STEM majors than students who did not. Murphy et al. (2010) investigated the impact of participating in a summer bridge program on student graduation rates. Such measures were significantly less common at the K–12 level. Assessment tools focused on engineering- and computing-related attitude and interest surveys were primarily used instead (Ericson et al., 2014; Gardner-McCune et al., 2013; Ladeji-Osias et al., 2016). For example, Oliveira (2010) investigated the impact of a summer STEM course on student attitudes toward college and STEM, also measuring student competence in electronics and circuit design.

The reporting practices varied depending on whether the instruments used were self-constructed by the practitioners conducting the assessment or identified from the literature. Of the articles that utilized survey instruments, many did not include discussions of the validity and reliability of the scales being used (Clark et al., 2015; Donnelly, 2007; Ericson et al., 2014; Ferreira, 2002), likely due to BPE stakeholders using self-constructed instruments
to account for program specifics. For example, Donnelly (2007) constructed survey items specifically related to program components, such as monthly meetings or the computer and teaching training they received to measure how agreeable the students were to the fact that these components were beneficial to their development. For the articles using established questionnaires, aspects of validity or reliability were reported in the text or pointed to previously reported literature (Di Salvo et al., 2013; Dunn & Veltman, 1989; D. Knight et al., 2013; B. Martin, 2014; Moyer, 2013). For example, Di Salvo et al. (2013) used the Microcomputer Beliefs Inventory (Enochs et al., 1993) and the Computer Attitude Questionnaire (Christensen & Knezek, 2001); their assessment reports included the scales’ reliability ratings (according to their Cronbach's alpha scores). Though validity and reliability are more often discussed in educational research, these constructs are often important when discussing the quality of assessment. As previously noted, the only way for BPE practitioners to adequately monitor progress is for them to be aware of the important issues that surround assessment measures. Having confidence that your measurements are measuring what you think they are (i.e., are valid) and that the results they produce are consistent (i.e., reliable) is directly connected with your ability to make claims about the impact of an effort and/or determine the extent to which it confirms or denies previous literature.

6.5 Theme 5: The assessment articles in our study most frequently highlighted positive evaluative claims about program performance rather than assessing areas for change or assessing the impact of those changes

Assessment results were typically used to highlight successful aspects of a program or provide evidence of its impact or effectiveness (Anderson-Rowland & Ruben, 2008; E. Brown, 2007; Clark et al., 2015; Griggs et al., 2016; Oliveira, 2010). For example, across articles, BPE stakeholders assert that their K–12 programs are effective based on an increase in positive attitudes and perceptions about STEM/CS, intent to pursue college, and intent to major in a STEM-related field (Ferreira, 2002; Ladeji-Osias et al., 2016). They similarly assert that their undergraduate programs are effective based on student retention, student achievement of internships, and graduate placement (Kendricks & Arment, 2011; Lam et al., 1997; McCullough et al., 2014). It should be noted that findings mostly served to justify why the said program was effective and should continue to be supported, rather than what could be done to improve the program.

Others were less matter of fact about the claims they made from their results and openly discussed their limitations and areas for improvement. Assessments focused on interventions for K–12 students expressed an understanding that other factors are simultaneously at play and, as a result, make it difficult to make strong claims about the results observed (Ericson et al., 2014). This challenge means that they can make few claims regarding the long-term impact of their interventions (Di Salvo et al., 2013). Authors also voiced a need for additional research and evaluation to make substantial claims about the outcomes of their programs (Dunn & Veltman, 1989; Ferreira, 2002; Murphy et al., 2010). Some program evaluators concluded articles with ideas for improving their intervention given the results of the assessment (Anderson-Rowland & Ruben, 2008; Ericson et al., 2014; Gardner-McCune et al., 2013). Moreover, Ferreira (2002) went a step further and made the argument that there is a need for education reform.

Given the pressures for BPE practitioners to prove their value, it is no surprise that not all authors highlighted places for improvement or any shortcomings that were revealed through their assessments. Similarly, knowing that resources are often scarce, the lack of discussions regarding the impact of changes that were made was not surprising. Just as it costs to host the initial intervention, it is also costly and time consuming to revisit and make changes to it.

7 IMPLICATIONS AND FUTURE WORK

Our research has important implications for the use of assessment and evaluation, and the assessment cycle, in educational practices related to BPE. Our paper sought to examine assessment/evaluation practices as they relate to BPE efforts. Although we have discussed a variety of topics, attributes, and considerations, we now offer implications grounded in the assessment cycle to assist practitioners with (1) planning and identifying outcomes; (2) collecting and analyzing data; (3) reporting and sharing results; and (4) identifying, implementing, and assessing change. Though our study was motivated by our research teams’ focus on Black Americans, our findings are broadly applicable to BPE stakeholders focused on a variety of underrepresented or otherwise marginalized groups. Future exploration should consider disaggregating data based on levels (i.e., individual, program, institution) given that the associated assessment
or evaluation may look differently as purposes differ and explore the extent to which BPE assessment/evaluation is implemented differently depending on the target population.

7.1 Implications for planning and identifying outcomes

First and foremost, practitioners should clearly state the questions, objectives, and/or goals when creating an assessment plan. Banta et al. (1996) note that assessment is most effective when it is based on a clearly stated purpose where goals and objectives are clearly understood. Suskie (2008) echoes this importance of assessment activities flowing directly from clearly understood goals. In doing so, it is important to focus not only on the success stories—as most assessment articles we reviewed sought to demonstrate how well their programs were preparing students—but also on deficiencies and areas for improvement. If we are to make progress in this area, we must be willing to critically review existing efforts and constantly share ways to improve. Doing so requires establishing a culture of trust, where practitioners do not fear that their programs' sustainability will be jeopardized by sharing results about things that did not go well.

Practitioners should also specify the level of assessment in their assessment reports. Doing so is important because this decision greatly informs the types of questions and objectives being set in an assessment plan (R. Miller & Leskes, 2005). As our results suggest, assessment or evaluation in the context of BPE is seldom used (or published) at the institution level; this gap in the literature presents an opportunity for future work and scholarship.

Lastly, there is a need for more publications that focus on efforts beyond the context of PWIs and the K-12 level. Program context matters when it comes to complex issues, and the lack of assessment publications focused on efforts in other settings, such as HBCUs and industry, has implications for our ability to address this challenge, particularly as it relates to Black Americans. If a large percentage of the population of interest attends a particular type of university, for example, it is important to capture what interventions do and do not work in those contexts. Additionally, there is a need for more work that focuses on analyzing the short- and long-term impact of interventions in various environments.

7.2 Implications for data collection and analysis

There is a need for well-developed and broadly shared data collection measures, such as validated survey instruments. Shared data collection measures are particularly important in the context of BPE because, though the assessments are conducted locally, the nation-wide nature of the challenge heightens the importance of being able to trust other's results. According to Singleton and Straits (2010b), surveys provide the most effective means of social description and can address a broad range of research topics. However, these instruments are only as good as their construction, and there are numerous factors one must consider during development—such as order and flow, question format, and the inclusion of bias in item questions (Singleton & Straits, 2010a). Robust qualitative data collection protocols are also needed to tell a more comprehensive story about the intervention and its impact on participants. Qualitative data bring rich context from the voices of those experiencing the program/intervention being assessed. We understand the time-consuming nature of qualitative data analysis that follows data collection; however, we urge BPE stakeholders to consider its value add.

There is also a need for more consistent reporting of results using quality metrics. The use of reliable and valid instruments increases the credibility of results tremendously, and reporting on the validity and reliability of the instruments used is just as important as selecting them. While some studies included validity and reliability measures, they were rare. Without such information, it is difficult for the reader to evaluate the credibility of the results and/or determine which results are transferable. If the internal validity is not confirmed, the external validity is compromised. When considering the indicators used to evaluate the quality of qualitative work (e.g., transparency, credibility, ethical considerations, and meaningful coherence) in educational research (Tracy, 2010), it should be noted that these indicators are applicable here as it relates to reporting on both qualitative and quantitative methods.

Lastly, there is also a need for more longitudinal designs that connect to the societal need for a more diverse engineering workforce. The absence of such designs creates a gap in understanding the long-term effectiveness of programs because many BPE interventions implicitly aim to contribute to the production of a more diverse
engineering workforce. Longitudinal designs can also aid in tracking changes across all levels of assessment, specifically at the institutional level. Given that some schools of thought surrounding change in higher education view change as slow, continuous, and long-term (Kezar, 2001), longitudinal designs aid in our assessment and evaluation at the institutional level.

7.3 | Implications for reporting and sharing results

Future assessment reports should be more transparent. Here, we learn from suggestions intended for educational research: “Transparent research is marked by [the] disclosure of the study’s challenges and unexpected twists and turns, and [the] revelation of the ways research foci transformed over time” (Tracy, 2010, p. 842). We argue that the same could apply to assessment because one of the major barriers to change in assessment is a lack of trust; increased transparency could help strengthen trust between BPE practitioners and their stakeholders.

With transparency in mind, it may be harder to forget specific details about various aspects of the assessment process. For example, providing more clarity on the program structure in conjunction with discussions about the results is imperative. The easier it is for readers to understand the context and the findings, the more trustworthy a report will read. The use of properly labeled tables and figures reporting on participant demographics and program components (e.g., logic models) is also needed, as this was not included in many of the documents. Additionally, reports should document the participant selection criteria because this information allows for stakeholders to make better supported claims about student performance. Lastly, we reiterate that negative findings were rarely published. We encourage practitioners to share both their success and their shortcomings and lessons learned. According to the literature we reviewed, it would appear that almost everything is working effectively, which we know anecdotally is not the case. This reality leads us to also call upon administrators and other stakeholders to support and believe the reports and statements by BPE practitioners who discuss and advocate for the needs of their students. BPE practitioners need this support from stakeholders to ensure that their work is valued and uplifted so that they can be transparent without the fear of repercussions (i.e., loss of job, demotion) for reporting on failures, shortcomings, and aspects that require improvement, change, or reform.

Lastly, when reporting and sharing results, practitioners should also be mindful of narratives they are creating in the process (i.e., deficit vs. asset-based). The framing of these results can impact how minoritized communities are represented within engineering and thus impact the types of efforts created to address BPE-related issues. Practitioners should be particularly mindful of this concern when planning to use student achievement as data metrics.

7.4 | Implications for identifying, implementing, and assessing change

Future assessment/evaluation reports should document insights related to all parts of the assessment cycle. We acknowledge that attending to the end of the assessment cycle is the hardest to complete (Kuh et al., 2015). As a result, it is not surprising that the implications of the assessment results and future work were seldom mentioned. It is also not surprising that more thorough discussions surrounding the end of the assessment cycle in BPE were seldom found in our literature search. Unfortunately, this significantly limits our ability to learn from the work done by others. While this omission does not mean that assessment/evaluation results did not inform changes, the absence of this information may leave a reader wondering about the utility of the findings—for those hosting the intervention and stakeholders desiring to learn from them.

Finally, it is worth acknowledging the challenges associated with documenting, tracing, and making sense of the impact of an intervention, regardless of the context. J. London and Cox (2015) note that these challenges center around issues of attribution, assessment/evaluation, and interpretation. The inherent challenges associated with demonstrating impact in any context simply add even more pressure when engaging in impact-related exercises in the BPE context. Until the issues associated with impact, in general, are resolved, BPE practitioners can only do so much regarding reporting when key stakeholders (especially donors or sponsors) desire absolute certainty of impact as the minimum starting place. Instead of focusing on executing high-quality interventions with fidelity, they are forced to focus more attention on gathering evidence to demonstrate value.
8 | CONCLUSION

The purpose of this study was to explore the presence of the assessment cycle in the literature related to BPE. In doing so, we hoped to illuminate the extent to which the entire assessment cycle is explored in the engineering education literature. We also intended to provide an overview of the nuances associated with planning, enacting, and using assessment/evaluation in educational practices. By grounding our SLR of assessment articles in a six-stage assessment cycle, we were able to not only highlight aspects of the cycle that were present but also emphasize that which was missing. For example, we note the omission of negative results as well as information about the change that resulted from the process. This bias toward positive reporting in the articles we reviewed is of concern because we have to be comfortable saying “This did not work” if we are going to move forward. We argue that the constant and overwhelming pressures to justify the need for and use of resources or demonstrate efficiency and ROI make doing valuable and effective work harder for BPE practitioners.

The uneven amount of pressure applied to BPE efforts must also be revisited if we (as a field) agree that broadening participation of historically underrepresented ethnic/racial groups is important. Moreover, there is room to strengthen the trust between BPE practitioners and other stakeholders. BPE practitioners need to receive unwavering support from stakeholders that have the power to reduce (ideally eliminate) the fear of repercussions (i.e., loss of job, demotion) to be transparent about failures and aspects that are problematic. Ultimately, there is room to improve how assessment information is reported, particularly as it relates to BPE.

ACKNOWLEDGMENTS

This material is based on work supported by the National Science Foundation under Award Numbers 1647327 and 1926935. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We thank Amy Van Epps for her guidance in conducting this work.

ORCID

Teirra K. Holloman https://orcid.org/0000-0001-8699-2466
Walter C. Lee https://orcid.org/0000-0001-5082-1411

REFERENCES

References marked with an asterisk (*) indicate studies included in the SLR.

Allen, M., Noel, R. C., Rienzi, B. M., & McMillin, D. J. (2002). Outcomes assessment handbook. California State University, Institute for Teaching and Learning.

*Anderson-Rowland, M., & Ruben, C. (2008). Academic achievement and retention in a minority engineering program. American Society for Engineering Education.

Baber, L. D. (2015). Considering the interest-convergence dilemma in STEM education. The Review of Higher Education, 38(2), 251–270. https://doi.org/10.1353/rhe.2015.0004

Bachiri, Z. (2019). Pathways: Engaging tech workers in the career paths of underserved youth. Carnegie Mellon University.

Banta, T. W., Lund, J. P., Black, K. E., & Oblander, P. W. (1996). Assessment in practice: Putting principles to work on college campuses. Jossey-Bass.

Borrego, M., Foster, M. J., & Froyd, J. E. (2014). Systematic literature reviews in engineering education and other developing interdisciplinary fields. Journal of Engineering Education, 103(1), 45–76. https://doi.org/10.1002/jee.20038

*Brown, E. (2007). The ST@R project: An initiative to increase the retention rates of 1st and 2nd year underrepresented students enrolled in electrical engineering. Paper presented at the ASEE Annual Conference and Exposition. Retrieved from https://peer.asee.org/3031

Brown, J. T. (2017). The seven silos of accountability in higher education: Systematizing multiple logics and fields. Research and Practice in Assessment, 11(Anniversary Issue), 41–58.

Chen, D. A., Mejia, J. A., & Breslin, S. (2019). Navigating equity work in engineering: Contradicting messages encountered by minority faculty. Digital Creativity, 30(4), 329–344. https://doi.org/10.1080/14626268.2019.1678486

Christensen, R., & Knezek, G. (2001). Instruments for assessing the impact of technology in education. Computers in the Schools, 18(2–3), 5–25. https://doi.org/10.1300/J025v18n02_02

*Clark, C. J., Ardley, T. W., & Black, J. T. (2015). The program of excellence in STEM: Involvement of traditionally underrepresented students in STEM education through research and mentoring at Florida A&M University. Paper presented at the 2015 IEEE Integrated STEM Education Conference, Princeton, NJ. https://doi.org/10.1109/ISECon.2015.7119915
Cook, D. J., Mulrow, C. D., & Haynes, R. B. (1997). Systematic reviews: Synthesis of best evidence for clinical decisions. *Annals of Internal Medicine, 126*(5), 376–380. https://doi.org/10.7326/0003-4819-126-5-19970310-00006

Cooper, M., Bullard, L. G., Spencer, D., & Willis, C. (2019). Direct and indirect assessment of student perspectives and performance in an online/distance education chemical engineering bridging course. Paper presented at the ASEE Annual Conference and Exposition, Tampa, FL. Retrieved from https://peer.asee.org/32665

*Di Salvo, B., Guzdial, M., Meadows, C., McKlin, T., Perry, K., & Bruckman, A. (2013). Workifying games: Successfully engaging African American gamers with computer science. Paper presented at the SIGCSE’13: Proceedings of the 44th ACM Technical Symposium on Computer Science Education, Denver, CO. https://doi.org/10.1145/2445196.2445292

*Donnelly, A. E. (2007). The south east alliance for graduate education and the professoriate program: Graduate minority retention and preparedness for academic careers. Paper presented at the ASEE Annual Conference and Exposition. Retrieved from https://peer.asee.org/2466

*Dunn, C. W., & Veltman, G. C. (1989). Addressing the restrictive career maturity patterns of minority youth: A program evaluation. *Journal of Multicultural Counseling and Development, 17*(4), 156–164. https://doi.org/10.1002/j.2161-1912.1989.tb00428.x

Enochs, L. G., Riggs, I. M., & Sc Ellis, J. D. (1993). The development and partial validation of microcomputer utilization in teaching efficacy beliefs instrument in a science setting. *School Science and Mathematics, 93*(5), 257–263.

*Ericson, B., Engelman, S., McKlin, T., & Taylor, J. (2014). Project rise Up 4 CS: Increasing the number of Black students who pass advanced placement CS A. Paper presented at the 45th ACM Technical Symposium on Computer Science Education, Atlanta, GA. https://doi.org/10.1145/2538862.2538937

Felder, R. M., & Brent, R. (2003). Designing and teaching courses for the ABET engineering criteria. *Journal of Engineering Education, 92*(1), 7–25.

*Ferreira, M. (2002). Ameliorating equity in science, mathematics, and engineering: A case study of an after-school science program. *Equity & Excellence in Education, 35*(1), 43–49. https://doi.org/10.1080/073845242

Finelli, C. J., Borrego, M., & Rasoulifar, G. (2015). Development of a taxonomy of keywords for engineering education research. *Journal of Engineering Education, 104*(4), 365–387. https://doi.org/10.1002/jee.20101

Finelli, C. J., & Froyd, J. E. (2019, Spring). Advances in engineering education improving student learning in undergraduate engineering education by improving teaching and assessment. *Advances in Engineering Education, 1*, 3–30.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program evaluation: Alternative approaches and practical guidelines* (4th ed.). Pearson Education.

Fletcher, T., Ross, M., Tolbert, D., Holly, J., Cardella, M., Godwin, A., Deboer, J., & Yates, N. (2017). Ignored potential: A collaborative road map for increasing African-American women in engineering. The National Society of Black Engineers, The Society of Women Engineers, and Women in Engineering ProActive Network.

*Gardner-McCune, C., McCune, D. B., Edwards, C. M., & Stallworth, C. (2013). I-3 experience: Expanding research and design opportunities for underrepresented high school students. Paper presented at the ASEE Annual Conference and Exposition. Retrieved from https://peer.asee.org/19691

Geiger, R. L. (2005). The ten generations of higher education. In P. G. Altbach, P. J. Gumport, & R. O. Berdahl (Eds.), *American higher education in the twenty-first century* (3rd ed., pp. 37–68). Johns Hopkins University Press.

George, C. E., Castro, E. L., & Rincon, B. (2019). Investigating the origins of STEM intervention programs: An isomorphic analysis. *Studies in Higher Education, 44*(9), 1645–1661. https://doi.org/10.1080/03075079.2018.1458224

Gibbons, M. T. (2009). *Engineering by the numbers: 2008–2009*. American Society for Engineering Education.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report, 8*(4), 597–607.

*Good, J., Halpin, G., & Halpin, G. (2002). Retaining Black students in engineering: Do minority programs have a longitudinal impact? *Journal of College Student Retention: Research, Theory and Practice, 3*(4), 351–364. https://doi.org/10.2307/2696252

*Good, J. M., Halpin, G., & Halpin, G. (2000). A promising prospect for minority retention: Students becoming peer mentors. *Journal of Negro Education, 69*(4), 375–383. https://doi.org/10.2307/2696252

*Griggs, L., Stringer, J. K., Rankins, F., & Hargraves, R. H. (2016). Investigating the impact of a hybrid summer transition program. Paper presented at the IEEE Frontiers in Education Conference, Rapid City, SD. https://doi.org/10.1109/FIE.2016.7757527

*Hanflin, L., & Schumack, M. (2001). *The Ford/University of Detroit mercy engineering opportunity program*. Paper presented at the ASEE Annual Conference and Exposition, Albuquerque, NM. Retrieved from https://peer.asee.org/9280

Hansen, E. J. (2012). Assessment, Part 1: Educative assessment. In *Hanifin, L., & Schumack, M. (2001).*

Higgins, J. P. T., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions (version 5)* (Vol. 5). The Cochrane Collaboration. Retrieved from https://www.cochrane-handbook.org

Hoej, J. J., & Nault, E. W. (2008). Barriers and challenges to assessment in engineering education. In J. E. Spurlin, S. A. Rajala, & J. P. Lavelle (Eds.), *Designing better engineering education through assessment: A practical resource for faculty and department chairs on using assessment and ABET criteria to improve student performance* (pp. 171–189). Stylus Publishing.

Holloman, T. K., Lee, W. C., London, J. S., Hakiyo, A. B., Jew, G., & Watford, B. A. (2018). A historical and policy perspective on broadening participation in STEM: Insights from national reports (1974–2016). Paper presented at the Collaborative Network for Engineering and Computing Diversity Conference, Crystal City, VA.

*Hrabowski, F. A., & Maton, K. I. (1995). Enhancing the success of African-American students in the sciences: Freshman year outcomes. *School Science & Mathematics, 95*, 19-27. https://doi.org/10.1177/00376421969631
Moskal, B., Leydens, J., & Pavelich, M. (2002). Validity, reliability and the assessment of engineering education. *Journal of Engineering Education, 91*(3), 351–354.

Moyer III, J. F. (2013). Probeware in 8th grade science: A quasi-experimental study on attitude and achievement. (Publishing No. 3552686.) [Doctoral dissertation, Wilmington University]. ProQuest Dissertations Publishing.

Mullen, P. D., & Ramirez, G. (2006). The promise and pitfalls of systematic reviews. *Annu. Rev. Public Health, 27*, 81–102.

*Murphy, T. E., Gaughan, M., Hume, R., & Moore, S. G. J. (2010). College graduation rates for minority students in a selective technical university: Will participation in a summer bridge program contribute to success? *Educational Evaluation and Policy Analysis, 32*(1), 70–83. https://doi.org/10.3102/0162373709360064

National Academy of Engineering. (2014). *Surmounting the barriers: Ethnic diversity in engineering education: Summary of a workshop*. National Academies Press.

National Research Council. (2001). In J. Pelligrino, N. Chudowsky, & R. Glaser (Eds.), *Knowing what students know: The science and design of educational assessment*. National Academies Press. https://doi.org/10.17226/10019

*Ogilvie, A. (2007). EOE first year interest groups: A success model for increasing retention. Paper presented at the ASEE Annual Conference and Exposition.

Olds, B. M., Moskal, B. M., & Miller, R. L. (2005). Assessment in engineering education: Evolution, approaches and future collaborations. *Journal of Engineering Education, 94*(1), 13–25. https://doi.org/10.1002/j.2168-9830.2005.tb00826.x

*Oliveira, A. (2010). *Impacting students' interest in stem fields: An electronic communication course for K–12 underrepresented students*. Paper presented at the ASEE Annual Conference and Exposition, Louisville, KY. Retrieved from https://peer.asee.org/15714

Palomba, C. A., & Banta, T. W. (1999). *Assessment essentials: Planning, implementing, and improving assessment in higher education*. Jossey-Bass.

Patton, L. D., Sánchez, B., Mac, J., & Stewart, D. L. (2019). An inconvenient truth about “progress”: An analysis of the promises and perils of research on campus diversity initiatives. *The Review of Higher Education, 42*(3), 173–198.

Pellegrino, J. W., DiBello, L. V., & Brophy, S. P. (2014). The science and design of assessment in engineering education. In A. Johri & B. M. Olds (Eds.), *Cambridge handbook of engineering education research* (pp. 571–598). Cambridge University Press.

Petitcrew, M., & Roberts, H. (2006). Systematic reviews in the social sciences: A practical guide. Blackwell Publishing.

Phillips, C. M. L., London, J. S., Lee, W. C., Epps, A. S. Van, & Watford, B. A. (2017). Reflections on the messiness of initiating a systematic literature review on broadening participation in engineering and computer science. Paper presented at the Frontiers in Education Conference, Indianapolis, IN.

Riley, D. (2016). *We assess what we value: “Evidence-based” logic and the abandonment of “non-assessable” learning outcomes*. Paper presented at the ASEE Annual Conference and Exposition, New Orleans, LA. https://doi.org/10.18260/p.27187

Rincón, B. E., & George-Jackson, C. E. (2016). Examining department climate for women in engineering: The role of STEM interventions. *Journal of College Student Development*, 57(6), 742–747. https://doi.org/10.1353/csd.2016.0072

Rodriguez, R. (2007). Life after Hopwood—University of Texas stops using standardized tests for admissions. *Diverse: Issues in Higher Education*. Retrieved from http://diverseeducation.com/article/7661/

Roy, J. (2019). *Engineering by the numbers: 2017–2018*. American Society for Engineering Education.

Shehab, R., Murphy, T. J., & Foor, C. E. (2012). “Do they even have that anymore”: The impact of redesigning a minority engineering program. *Journal of Women and Minorities in Science and Engineering, 18*(3), 235–253. https://doi.org/10.1615/JWomenMinorScienEng.2013002354

Singleton, R. A., & Straits, B. C. (2010a). Survey instrumentation. In *Approaches to social research* (pp. 309–351). Oxford University Press.

Singleton, R. A., & Straits, B. C. (2010b). Survey research. In *Approaches to social research* (pp. 263–307). Oxford University Press.

Stagner, J. A., & Jorendt, J. L. (2015). *Assessing graduate attributes within a two-semester capstone design course*. Paper presented at the ASEE Annual Conference and Exposition, Seattle, WA. Retrieved from https://peer.asee.org/23581

Suskie, L. (2008). Understanding the nature and purpose of assessment. In J. E. Spurlin, S. A. Rajala, & J. P. Lavelle (Eds.), *Designing better engineering education through assessment: A practical resource for faculty and department chairs on using assessment and ABET criteria to improve student learning* (pp. 3–19). Stylus Publishing.

*Taylor, O. L., McGowan, J., & Alston, S. T. (2008). The effect of learning communities on achievement in STEM fields for African Americans across four campuses. *Learning Communities and the Higher Education of African Americans, 77*(3), 190–202.

Thelin, J. R. (2011). A history of American higher education. John Hopkins University Press.

Town, C., & Pierce, W. (2013). *Glossary of assessment terms*. American Public University System. Retrieved from http://www.apus.edu/community-scholars/learning-outcomes-assessment/university-assessment/glossary.htm

Tracy, S. J. (2010). Qualitative quality: Eight “big-tent” criteria for higher qualitative research. *Qualitative Inquiry, 16*(10), 837–851. https://doi.org/10.1177/1077800410383121

*White, J. L., Altschuld, J. W., & Lee, Y.-F. (2008). Evaluating minority retention programs: Problems encountered and lessons learned from the Ohio science and engineering alliance. *Evaluation and Program Planning, 31*(3), 277–283. https://doi.org/10.1016/j.evalprogplan.2008.03.006

*Wilson, J. A., Gaines, J. E., & Cooper, D. (2015). Bulls-EYE mentoring: Developing a program intervention in the college of engineering. Paper presented at the ASEE Annual Conference and Exposition, Seattle, WA. Retrieved from https://peer.asee.org/23650

Winter, G. (2000). A comparative discussion of the notion of “validity” in qualitative and quantitative research. *The Qualitative Report, 4*(3), 1–14. Retrieved from http://www.nova.edu/ssss/QR/QR4-3/winter.html
AUTHOR BIOGRAPHIES

**Teirra K. Holloman** is a Doctoral Candidate in Engineering Education at Virginia Tech, 635 Prices Fork Road, Blacksburg, VA 24061; teirrah@vt.edu.

**Walter C. Lee** is an Associate Professor of Engineering Education and the Director for Research in the Center for the Enhancement of Engineering Diversity (CEED) at Virginia Tech, 635 Prices Fork Road, Blacksburg, VA 24061; walterl@vt.edu.

**Jeremi S. London** is an Assistant Professor of Engineering Education at Virginia Tech, 635 Prices Fork Road, Blacksburg, VA 24061; jslondon@vt.edu.

**Chaneé D. Hawkins Ash** is the Co-founder and Principal of Foresight Strategy Solutions, 2033 Illinois Street, Vallejo, CA 94590; chanee.ash@gmail.com.

**Bevlee A. Watford** is a Professor of Engineering Education, Associate Dean for Equity and Engagement, and Executive Director of the Center for the Enhancement of Engineering Diversity in the College of Engineering at Virginia Tech, 490 Old Turner Street, Blacksburg, VA 24061; deuce@vt.edu.

---

How to cite this article: Holloman, T. K., Lee, W. C., London, J. S., Hawkins Ash, C. D., & Watford, B. A. (2021). The assessment cycle: Insights from a systematic literature review on broadening participation in engineering and computer science. *Journal of Engineering Education, 110*(4), 1027–1048. [https://doi.org/10.1002/jee.20425](https://doi.org/10.1002/jee.20425)