



AUTHORS

Kristen L. Smith, M.A.
James Madison University

Megan Rodgers Good, M.A.
James Madison University

Elizabeth Hawk Sanchez, B.A.
James Madison University

Keston H. Fulcher, Ph.D.
James Madison University

Abstract

Although higher education institutions often engage in assessment practices, use of assessment results to improve student learning is rare (Blaich & Wise, 2011). We surmised that this rarity could be partially explained by unclear communication regarding what *use of results* means. The current study qualitatively investigated how assessment professionals define *use of assessment results to improve student learning* in assessment literature, assessment rubrics, and regional accreditation standards. We found that most definitions were vague and lacked detailed examples. This ambiguity may help explain why using results to make data-supported curricular or pedagogical changes and then re-assessing students to determine the effect of those changes is so uncommon in higher education. Based on our findings, we clarify what it means to *close the loop* in an effort to facilitate greater use of results to evidence improved student learning.

Communication is Key: Unpacking “Use of Assessment Results to Improve Student Learning”

Prior to the 1980s, external stakeholders evaluated the quality of U.S. colleges based on inputs and outputs such as average entrance scores, the number of books in a library, and graduation rates (Erwin, 1991). In 1985, higher education scholar Alexander Astin suggested that talent development was an alternative, better measure of quality. Similarly, Barr and Tagg (1995) called for a greater focus on outcomes, claiming that it would be more advantageous to fund an institution based on the number of math problems students solve rather than based on the number of students who sit in a math class. Barr and Tagg reiterated that higher education systems should reflect and fulfill their responsibilities to promote student learning. In other words, the emphasis should be on how much students have learned or developed as a function of the institution—not on how many students attended class.

Today, in an era of skepticism regarding the value of education, colleges and universities would benefit from demonstrating that student learning is improving. In fact, some have suggested the importance of learning improvement by calling it the bottom line of education. And, like businesses, institutions should endeavor to optimize their [learning] bottom line (Clarke, 2002). In the late 1980s, legislators crafted policy reflecting this re-imagining of quality (Ewell, 2009). From that point forward, institutions—under federal mandate—have been assessing and reporting on student learning outcomes. The idea was as follows: If institutions carefully defined student learning outcomes and assessed them, they would be well positioned to make changes that would enhance or improve student learning.

After 25 years of defining and assessing student learning outcomes, one would presume that many institutions could evidence improved student learning. Unfortunately, such evidence has not proliferated as quickly as the practice of assessment itself. In 1996,

CORRESPONDENCE

Email
smith4kl@jmu.edu

After 25 years of defining and assessing student learning outcomes, one would presume that many institutions could evidence improved student learning. Unfortunately, such evidence has not proliferated as quickly as the practice of assessment itself.

Trudy Banta and colleagues provided a few cases of improved student learning associated with assessment practice, but they conceded that such cases were rare. More recently, after interpreting findings from a multi-university study on assessment and learning, Kuh (2011) concluded:

...most colleges and universities were using multiple measures to determine student learning outcomes. At the same time, relatively few schools were ‘closing the loop,’ or using the information in any material way to intentionally modify policy and practice. Rarer still were colleges or universities where changes in policies or practices made a positive difference in student attainment. (p. 4)

Given this state of affairs, one may ask why student learning outcomes assessment practice is so unsuccessful relative to its stated goal of improvement. A complete answer to that question is complex and beyond the scope of this study. Nevertheless, one culprit is unclear communication.

Ambiguous and Inconsistent Communication

Indeed, we have noticed that use of results is emphasized ubiquitously at assessment conferences, but it is not always clear what *use* means. For instance, we have commonly found assessment practitioners using vague terms such as *use of results*, *closing the loop*, *improvement*, *action plan*, and so forth to define *using results to evidence improved learning*. In such cases, terms are rarely explicated; meaning practitioners must subjectively interpret what these terms mean. Without clearly delineating this critical step in the assessment process, it is no wonder examples of learning improvement are so scarce. Perhaps it is time for assessment practitioners to abandon the ambiguous use of results terminology in favor of a more concrete, consistent definition of what it means to use assessment results for learning improvement.

Use of results for learning improvement has been defined as programs making a pedagogical or curricular change. However, as Fulcher, Good, Coleman, and Smith (2014) note, a change is not an improvement. Rather, use of assessment results should be defined in terms of strong evidence, from direct measures and re-assessment, supporting substantive student learning improvement due to program modifications. Fulcher and colleagues further explain that practitioners often make statements like, “We made x, y, and z improvements to the program,” when they really mean, “We made x, y, and z changes.” A change is an improvement only after one reassesses and actually demonstrates a positive effect on student learning.

The Need for Clearer Communication

Certainly, the ambiguous and inconsistent language used to describe use of results is crippling our ability, as assessment practitioners, to demonstrate improved student learning. We need a better way to communicate what it means to effectively close the assessment loop and to demonstrate that assessment results influenced improvements in student learning. To this end, authors, practitioners, accreditors, and other stakeholders must engage in purposeful discourse to clarify the language we use in daily conversations, at conference presentations, and in assessment resources. Furthermore, higher education professionals would benefit from having applied examples of demonstrably improving student learning at the academic program level. Such examples should be situated within contexts that are salient to higher education practitioners.

Investigating Common Definitions of Use of Results

Assessment literature and other academic resources typically describe the steps of the assessment process, including use of results (e.g., Erwin, 1991; Walvoord, 2010). Higher education professionals may seek information about use of results from multiple sources including assessment books, meta-assessment rubrics, and accreditation standards. Many assessment books are easily accessible and designed for practitioners just beginning their assessment work. At institutions where assessment practice is more mature, assessment

A change is an improvement only after one reassesses and actually demonstrates a positive effect on student learning.

practitioners may employ meta-assessment rubrics to measure the quality of programmatic assessment processes across an institution's programs (Fulcher & Orem, 2010). Meta-assessment rubrics typically define quality practices at each stage in the assessment cycle, including use of results. Lastly, given most practitioners must assess student learning outcomes and report the results to an accrediting organization, various regional accreditation standards are prevalent resources that describe assessment processes.

In the current study, we reviewed a selection of assessment books, a selection of meta-assessment rubrics, and the accreditation standards for six regional accrediting organizations to determine how authors defined use of results in reference to student learning improvement. More specifically, we rated each resource using five dichotomous (i.e., yes or no) criteria:

- a) directly references student learning or development;
- b) general mention of "use of results," "closing the loop," "improvement," etc.;
- c) mention or description of a change to curriculum;
- d) mention or description of a change to pedagogy or teaching;
- e) mention or description of the need to "re-assess" or determine whether "changes" contributed to actual "improvements."

In addition to the presence or absence of the five aforementioned criteria, we rated each resource in terms of *Level Specificity* and *Intervention Specificity* using a five-point scale (i.e., 0, 0.5, 1, 1.5, 2). Level Specificity referred to the degree to which use of results was defined at the program, department, or unit level. Intervention Specificity represented the degree to which each resource provided a detailed, real-life example of what is meant by use of results to improve student learning (see Figure 1).

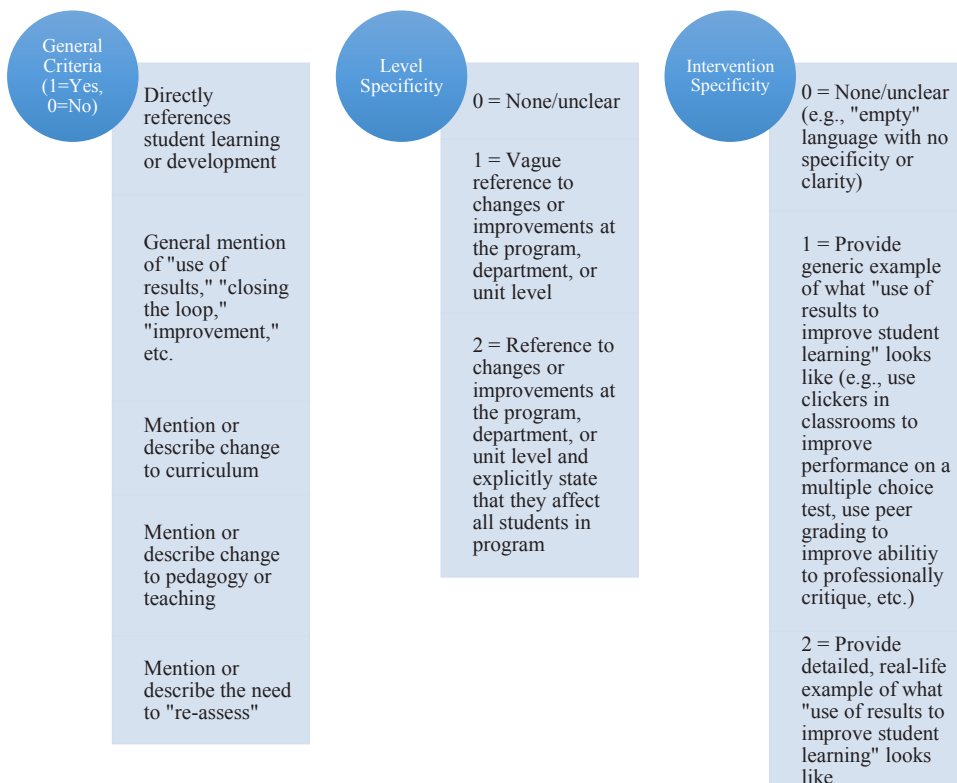


Figure 1. Criteria used to rate definitions of "use of results."

Perhaps it is time for assessment practitioners to abandon the ambiguous use of results terminology in favor of a more concrete, consistent definition of what it means to use assessment results for learning improvement.

A change is an improvement only after one reassesses and actually demonstrates a positive effect on student learning.

Data sources. We were interested in identifying assessment resources that an assessment novice might easily access via a limited Internet search. Thus, we initially used Google.com and Amazon.com to determine popular higher education assessment books that resulted from these searches (i.e., popular books meaning books displayed at the top of the list of search results that Google and Amazon provided). We conducted the Google.com and Amazon.com searches in October of 2013 using a non-personal, on-campus computer with a university IP address. Note, the cache was not cleared prior to conducting the search. More specifically, we searched the following terms: *higher education assessment*, *higher education assessment books*, and *student learning outcomes assessment books*.

Our initial search yielded discipline-specific resources and a number of books focused on classroom-level assessment, in addition to a few of the most popular, general assessment books (i.e., Bresciani, Gardner, & Hickmott, 2009; Suskie, 2010; Walvoord, 2010). Because we were interested in general program assessment resources we continued our search by selecting specific assessment books based on our collective assessment expertise.

We identified 14 higher education assessment books, and then we used the Table of Contents to identify the most relevant sections of each book pertaining to the use of assessment results. Thus, it is possible that each book could detail the use of assessment results in other sections; however, our approach considered that a practitioner would likely seek information about using results in the section of the book where this issue is highlighted (i.e., there is a section in the Table of Contents dedicated to use of results). Furthermore, we acknowledge that the 14 books we reviewed represented only a subset of all available assessment resource books.

To locate meta-assessment rubrics, we attempted to access the 58 rubrics identified by Fulcher, Swain, and Orem (2012). Unfortunately, many of the web links to these rubrics were no longer active and we were only able to locate 32 of the 58 meta-assessment rubrics. Thus, the 32 rubrics represented only a subset of all possible meta-assessment rubrics used across various higher education institutions. We only evaluated institutional meta-assessment rubrics (i.e., rubrics used at an institutional level to evaluate all academic programs, which may include both pre-professional and non-professional academic programs).

In the current study, we reviewed a selection of assessment books, a selection of meta-assessment rubrics, and the accreditation standards for six regional accrediting organizations to determine how authors defined use of results in reference to student learning improvement.

The majority (78.1%) of the 32 meta-assessment rubrics came from 4-year, public institutions. Also, nearly half (43.8%) were located in the North Central Association of Colleges and Schools accreditation region. Eight of the 32 institutions were classified as small (having fewer than 5,000 students), while 12 were medium (having 5,001-15,000 students), and the remaining 12 were large (having more than 15,000 students). Of the 32 meta-assessment rubrics we were able to locate online, none came from institutions located in the New England region. Therefore, none of the institutional meta-assessment rubrics we rated are from schools accredited by the New England Association of Colleges and Schools.

Lastly, we reviewed the standards for the six regional accreditors in the United States: Middle States Commission on Higher Education (2011), North Central Association of Colleges and Schools: The Higher Learning Commission (2014), New England Association of Schools and Colleges (2011), Northwest Commission on Colleges and Universities (2010), Southern Association of Colleges and Schools Commission on Colleges (2012), and Western Association of Schools and Colleges: Senior College and University Commission (2013). For each accrediting organization, we rated the specific section of their standards that related to assessment of student learning.

Procedures. Three of the authors of this article independently evaluated the 14 selected assessment books, 32 meta-assessment rubrics, and six regional accreditation standards. These three raters have extensive doctoral training in assessment, diverse experiences rating academic program assessment reports, and collectively 23 years of assessment consultation experience. In addition, the three raters helped create the rubric used to evaluate the assessment resources; therefore, they were familiar with the rubric and how to apply the various rubric components.

Table 1 displays the percent exact agreement for the five dichotomously rated criteria prior to rater adjudication. Collapsing across all five criteria, the average percent exact agreement was weakest for the books (83%) compared to the rubrics (90%) and accreditation standards

Table 1
Percent Exact Agreement for Five Criteria Prior to Rater Adjudication

| Rater | Directly references student learning or development | General mention of “use of results,” “closing the loop,” “improvement,” etc. | Mention or describe change to curriculum | Mention or describe change to pedagogy or teaching | Mention or describe the need to “re-assess” or determine whether “changes” contributed to actual “improvements” |
|---|---|--|--|--|---|
| Books (<i>N</i> = 14) | | | | | |
| 1 vs. 2 | 100 | 86 | 71 | 64 | 93 |
| 1 vs. 3 | 100 | 86 | 64 | 71 | 71 |
| 2 vs. 3 | 100 | 100 | 79 | 79 | 79 |
| Average % agreement | 100.0 | 90.7 | 71.3 | 71.3 | 81.0 |
| Rubrics (<i>N</i> = 32) | | | | | |
| 1 vs. 2 | 97 | 81 | 81 | 100 | 88 |
| 1 vs. 3 | 100 | 78 | 94 | 94 | 88 |
| 2 vs. 3 | 97 | 91 | 81 | 94 | 91 |
| Average % agreement | 98.0 | 83.3 | 85.3 | 96.0 | 89.0 |
| Accreditation Standards (<i>N</i> = 6) | | | | | |
| 1 vs. 2 | 100 | 100 | 100 | 100 | 100 |
| 1 vs. 3 | 100 | 100 | 100 | 100 | 100 |
| 2 vs. 3 | 100 | 100 | 100 | 100 | 100 |
| Average % agreement | 100 | 100 | 100 | 100 | 100 |

**Note.* Values are percent exact agreement. That is, if Rater 1 and Rater 2 had exact agreement for 13 out of the 14 books they rated on the “Directly references student learning or development” criteria, the percent agreement would be equal to 93% or 13 divided by 14. *Average % agreement* is the average percent exact agreement calculated for each of the five criteria for each rater pair.

(100%). This is likely because the books had far more information to be evaluated than the rubrics or standards. The *mention or describe a change to curriculum* criterion had the weakest average percent exact agreement (86%) compared to the other four dichotomously rated criteria. The average percent exact agreement was 91% across all five dichotomously rated criteria and resources.

After independently rating the five dichotomous criteria, raters adjudicated any discrepancies to reach exact agreement. Raters also adjudicated discrepancies for the Level Specificity and Intervention Specificity ratings. Given Level Specificity and Intervention Specificity were rated using a five-point scale, raters adjudicated to reach agreement within 0.5 points (i.e., ratings on a given criterion from two different raters must be within 0.5 points). Specifically, if two raters provided ratings on the same criterion that differed by more than 0.5 points, then raters engaged in a discussion of this discrepancy by providing a rationale to support or explain how they rated that specific criterion. The raters continued to discuss their ratings and explanations pertaining to a given criterion until all raters could reach agreement within 0.5 points. In some cases, Rater 1 may have missed information or a specific explanation, and Rater 2 subsequently identified where it could be found within the resource. Once Rater 1 saw the information or explanation she had missed during her independent rating, she typically agreed with Rater 2 and they easily adjudicated their ratings. In other cases, two raters might have interpreted text or information within the resource differently and a discussion ensued until the two raters achieved agreement within 0.5 points.

Because we created the rubric prior to evaluation and we had not used the rubric in previous research studies, there were exactly two instances in which we had to establish an additional adjudication rule. As described in the following paragraphs, we instated these two rules during the adjudication process after having in-depth conversations and agreement

To demonstrably improve student learning at a programmatic level, use of results should be explicitly and clearly defined in terms of the program, department, or unit level.

Nevertheless, meta-assessment rubrics and accreditation standards would likely be more helpful and valuable for practitioners if they included clear, explicit examples of use of results to improve student learning.

across all three raters about the rules. We aimed to achieve accurate ratings and ensure agreement on those ratings across raters. Furthermore, the two rules were mechanisms to clarify some of the language used in the rubric, not to tailor the rubric to the behaviors of the three raters. Clarifying these two aspects of the rubric during our adjudication processes was beneficial because it helped us apply the rubric in a more consistent and accurate way, and will help us do the same in future research studies. Additionally, if other faculty members want to use this rubric as part of future studies, we can use these two rules to help them understand the meaning of specific rubric criteria.

We noticed consistent disagreement between Rater 1 and Rater 2 on the Level Specificity criterion during adjudication; therefore, we created a rule to further clarify how to interpret this criterion: If the book, rubric, or accreditation standard mentioned anything that indicated the program level (e.g., unit, department) it would receive a rating of 2. During adjudication, we also realized that Raters 2 and 3 were interpreting the second criterion (general mention of “use of results,” “closing the loop,” “improvement,” etc.) more liberally, giving credit for terms such as action plan, while Rater 1 was only giving credit for the terms explicitly listed in the criterion (i.e., use of results, closing the loop, and improvement). Therefore, we created a rule that additional terms not mentioned in the criterion, such as action plan, would receive credit for the second criterion.

We calculated Cohen’s (1960) kappa to provide a more conservative estimate of inter-rater agreement. Kappa compares the agreement between two raters on a given criterion (e.g., agreement of Raters 1 and 2 on the directly references student learning or development criterion), taking into account chance agreement. The typical kappa value across all rater pairs and criteria, for all resources evaluated, was 0.486 with notable variability (i.e., values ranged from -0.148 to 1.000). The lowest kappa value was between Rater 1 and Rater 2 on the Level Specificity criterion. The kappa values for the second criterion (general mention of “use of results,” “closing the loop,” “improvement,” etc.) were also low. These lower kappa values were expected given the disagreements noted in the previous paragraph.

After revisiting our ratings during adjudication using these two rules, we felt confident in our ratings for the Level Specificity and the General Mention criteria. We also noted that a restriction of range could explain the lower kappa values. Although the Level Specificity and Intervention Specificity criteria were rated using a five-point scale (i.e., 0, 0.5, 1, 1.5, 2), only a few resources actually received a rating of 1.5 or 2.

Findings

Five Dichotomous Criteria

Books. Twelve of the 14 books directly referenced student learning or development, and generally mentioned use of results, closing the loop, improvement, and so forth (i.e., criteria 1 and 2). However, fewer books met the third (i.e., mention or description of a change to curriculum) and fourth criteria (i.e., mention or description of a change to pedagogy or teaching). Only three of the 14 books, Bresciani et al. (2009), Suskie (2010), and Walvoord (2010), mentioned or described the need to re-assess or determine if changes based on assessment results contributed to actual improvements. Results are reported in Table 2.

Meta-assessment rubrics. Note, we present the results for the 32 meta-assessment rubrics in aggregate form to preserve institutions’ confidentiality. It was impressive that the 32 institutions used meta-assessment rubrics because this requires mature assessment processes and adequate assessment infrastructure. Moreover, all of the meta-assessment rubrics directly referenced student learning or development, and generally mentioned use of results, closing the loop, improvement, and so forth. Approximately 38% of the rubrics defined use of results in terms of a change to curriculum, while 25% defined use of results in terms of a change to pedagogy or teaching (see Figure 2). Overall, the percentage of meta-assessment rubrics and the percentage of assessment books that mentioned the need to re-assess or determine whether changes based on assessment results contributed to actual improvements were comparable (i.e., 19% and 21%, respectively).

Overall, the assessment books provided some of the richest examples of using results for learning improvement. However, these examples were not the focus of the chapters in which they were found; thus, they may be difficult for readers to identify and internalize.

Table 2
Presence of the Five Criteria Used to Define “Use of Results” in Assessment Books

| Book Author (Publication Date) | Directly references student learning or development | General mention of “use of results,” “closing the loop,” “improvement”, etc. | Mention or describe change to curriculum | Mention or describe change to pedagogy or teaching | Mention or describe the need to “re-assess” or check to determine whether “changes” contributed to actual “improvements” |
|--|---|--|---|--|---|
| Allen (2004) | X | X | X | X | |
| Banta, Lund, Black, & Oblander (1996) | X | X | X | X | |
| Bresciani, Gardner, & Hickmott (2009) | X | X | X | X | X |
| Brown & Knight (1994) | X | X | X | X | |
| Erwin (1991) | X | X | | | |
| Huba & Freed (2000) | X | X | X | X | |
| Messick (1999) | X | X | | X | |
| Middaugh (2009) | X | X | | X | |
| Palomba & Banta (1999) | X | X | X | X | |
| Schuh (2009) | X | | | | |
| Schuh & Upcraft (2001) | X | X | | | |
| Suskie (2010) | X | X | X | X | X |
| Walvoord (2010) | X | X | X | X | X |
| Weiss (1998) | | X | X | X | |

*Note. X indicates the presence of the criterion (i.e., an X represents a rating of 1 for “Yes, the criterion was met”).

Figure 3 displays the number of meta-assessment rubrics (categorized according to regional accrediting organization) that met the third, fourth, and fifth criteria for defining use of results. A larger number of meta-assessment rubrics from institutions accredited by the Middle States Commission on Higher Education (MSCHE) met the fourth and fifth criteria compared to rubrics from institutions accredited by other regional accreditation organizations. However, similar to the findings for the assessment books, the fifth criterion (the need to re-assess) was the least frequently satisfied criterion for the meta-assessment rubrics.

Accreditation standards. As shown in Table 3, 100% of the regional accreditation standards directly referenced student learning or development, and generally mentioned use of results, closing the loop, improvement, and so forth. Only one of the regional accreditation standards (WASC) defined use of results in terms of a change to curriculum and a change to pedagogy or teaching. Interestingly, none of the regional standards defined use of results in terms of the need to re-assess or determine whether changes based on assessment results contributed to actual learning improvements.

Interestingly, none of the regional standards defined use of results in terms of the need to re-assess or determine whether changes based on assessment results contributed to actual learning improvements.

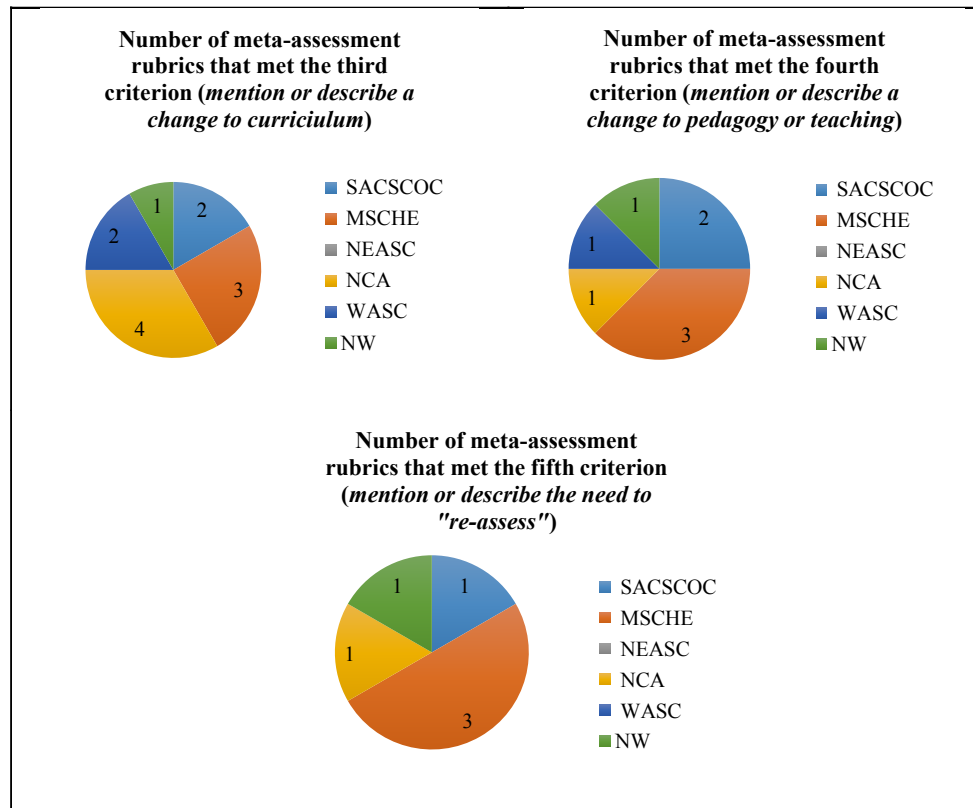
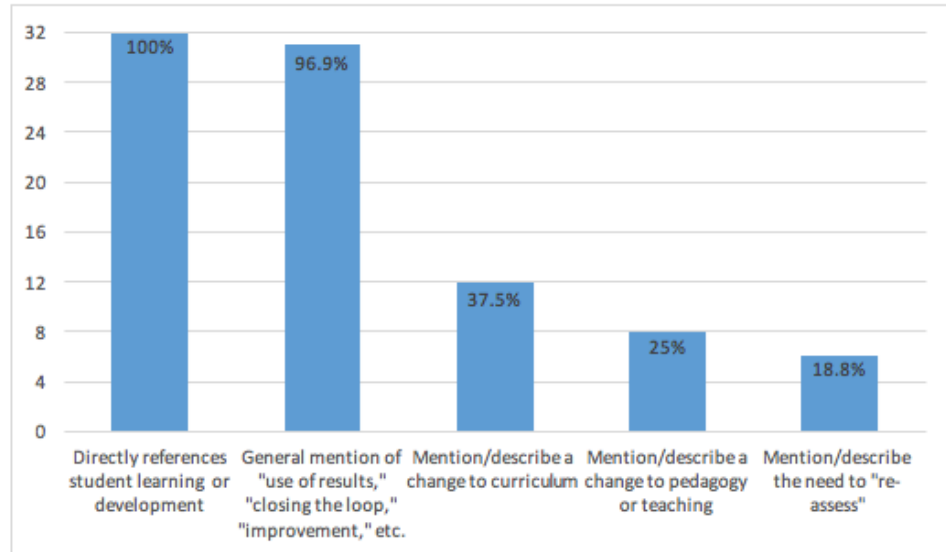


Figure 3. Number of meta-assessment rubrics that met the third, fourth, and fifth criteria for defining “use of results” categorized according to accrediting organization.

Level Specificity

To demonstrably improve student learning at a programmatic level, use of results should be explicitly and clearly defined in terms of the program, department, or unit level. Moreover, curricular or pedagogical modifications should affect every student completing the program (Fulcher et al., 2014). Thus, in addition to the presence or absence of the five criteria for defining use of results, we investigated the degree to which use of results was defined at the program, department, or unit level (Level Specificity).

Table 3
Presence of the Five Criteria Used to Define “Use of Results” in Regional Accreditation Standards

| Regional Accreditor | Directly references student learning or development | General mention of “use of results,” “closing the loop,” “improvement,” etc. | Mention or describe a change to curriculum | Mention or describe a change to pedagogy or teaching | Mention or describe the need to “re-assess” or check to determine whether “changes” contributed to actual “improvements” |
|---------------------|---|--|--|--|--|
| MSCHE | X | X | | X | |
| NCA | X | X | | | |
| NEASC | X | X | | | |
| NW | X | X | | | |
| SACSCOC | X | X | | | |
| WASC | X | X | X | X | |

**Note.* X indicates the presence of the criteria (i.e., an X represents a rating of 1 for “Yes, the criteria was met”). MSCHE = Middle States Commission on Higher Education, NCA = North Central Association of Colleges and Schools: The Higher Learning Commission, NEASC = New England Association of Schools and Colleges, NW = Northwest Commission on Colleges and Universities, SACSCOC = Southern Association of Colleges and Schools Commission on Colleges, WASC = Western Association of Schools and Colleges: Senior College and University Commission.

Level Specificity was evaluated using a five-point scale (0, 0.5, 1, 1.5, 2). The resources might not articulate any level when defining use of results to make changes or improvements. Or, it could be unclear what level is implicated in the definition (i.e., 0 = *None/unclear*). When defining use of results, the resources could vaguely reference the program level (i.e., 1 = *vague reference to changes or improvements at the program, department, or unit level*). However, an exemplary definition of use of results explicitly references changes or improvements that affect all students in a given program (i.e., 2 = *reference to changes or improvements at the program, department, or unit level and explicitly state that they affect all students in a program*). The meta-assessment rubrics received the highest ratings for Level Specificity compared to the books and accreditation standards (see Table 4). The accreditation standards were least clear about referencing use of results to make changes or improvements at the program level.

After reviewing and rating various assessment resources in search of exemplary definitions of use of results, we found that all had shortcomings in reference to evidencing learning improvement.

Table 4
Average Adjudicated “Level Specificity” Ratings for Books, Rubrics, and Standards

| | Average Adjudicated Rating |
|--|----------------------------|
| Books (<i>N</i> = 14) | 0.86 |
| Meta-assessment Rubrics (<i>N</i> = 32) | 1.00 |
| Accreditation Standards (<i>N</i> = 6) | 0.60 |

**Note.* Ratings on a five-point scale ranging from 0 to 2, with half points possible (i.e., 0, 0.5, 1, 1.5, 2). 0 = *None/unclear*; 1 = *vague reference to changes or improvements at the program, department, or unit level*; 2 = *reference to changes or improvements at the program, department, or unit level and explicitly state that they affect all students in a program*.

Intervention Specificity

In addition to defining use of results in terms of making changes or improvements at the program level, exemplary definitions should include an example. Research suggests that learning by example can facilitate conceptual understanding (Atkinson, Derry, Renkl, & Wortham, 2000; Bourne, Goldstein, & Link, 1964). Therefore, we investigated the degree to which resources provided an example of use of results to improve student learning (Intervention Specificity).

By situating use of results to improve student learning at the forefront of assessment practice, higher education professionals will likely experience greater buy-in to the assessment process and foster greater efforts to use assessment results in a meaningful way.

Intervention Specificity was evaluated using a five-point scale (0, 0.5, 1, 1.5, 2). That is, the resource may have provided no examples of using results to improve student learning or an example(s) that is vague or lacking sufficient detail (i.e., 0 = *None/unclear; “empty” language with no specificity or clarity*). Alternatively, the resources could have provided a generic example of what use of results to improve student learning means (i.e., 1 = *generic example; use clickers in classrooms to improve performance on a multiple choice test, use peer grading of capstone portfolios to improve ability to professionally critique, etc.*). Additionally, resources could have provided a detailed example of what use of results to improve student learning means (i.e., 2 = *detailed example; references assessment, modifications to pedagogy or curriculum, and re-assessment to determine whether modifications actually improved student learning*).

The assessment books had the highest ratings for Intervention Specificity compared to the meta-assessment rubrics and accreditation standards (see Table 5). Certainly, books might have a slight advantage in the Intervention Specificity category because books have more space to include examples of using assessment results to improve student learning. Nevertheless, meta-assessment rubrics and accreditation standards would likely be more helpful and valuable for practitioners if they included clear, explicit examples of use of results to improve student learning.

Table 5
Average Adjudicated “Intervention Specificity” Ratings for Books, Rubrics, and Standards

| | Average Adjudicated Rating |
|--------------------------------------|----------------------------|
| Books ($N = 14$) | 0.57 |
| Meta-assessment Rubrics ($N = 32$) | 0.08 |
| Accreditation Standards ($N = 6$) | 0.00 |

*Note. Ratings on a five-point scale ranging from 0 to 2, with half points possible (i.e., 0, 0.5, 1, 1.5, 2). 0 = *None/unclear; “empty” language with no specificity or clarity*; 1 = *generic example; use of clickers in classrooms to improve performance on a multiple choice test, use peer grading of capstone portfolios to improve ability to professionally critique, etc.*; and 2 = *detailed example; references assessment, modifications to pedagogy or curriculum, and re-assessment to determine whether modifications actually improved student learning*.

Discussion

Overall, the resources available to assessment practitioners did not clearly explicate use of results. However, we identified a few resources that had good definitions of use of results. These resources received some of the highest ratings across the five dichotomous criteria, as well as the Level Specificity and Intervention Specificity criteria.

Good Definitions of Use of Results

Assessment books. Banta, Lund, Black and Oblander’s (1996) *Assessment in Practice: Putting Principles to Work on College Campuses* was the only resource that received our highest rating (2 = *detailed example; references assessment, modifications to pedagogy or curriculum, and re-assessment to determine whether modifications actually improved student learning*) for Intervention Specificity. Banta and colleagues (1996) described examples from several institutions that provide “concrete evidence” of improved student learning (p. 343). Their examples included re-assessment as part of using results to evidence improved student learning. However, Banta and colleagues did not explicitly convey that re-assessment is part of how they defined use of results.

Barbara Walvoord's (2010) *Assessment Clear and Simple: A Practical Guide for Institutions, Departments, and General Education*, and Linda Suskie's (2010) *Assessing Student Learning: A Common Sense Guide* also received some of the highest ratings. Unlike many of the other resources, Walvoord described the need to follow-up or re-assess student learning after taking action. Furthermore, Walvoord's definition of using results appropriately differentiated classroom assessment from program-level assessment. Suskie's book received high ratings because she included general examples of what it means to use assessment results for improvement. These examples included re-assessing to verify that changes were indeed improvements. Suskie also discussed the importance of using results for pedagogical professional development among faculty and creating curricular coherence.

Overall, the assessment books provided some of the richest examples of using results for learning improvement. However, these examples were not the focus of the chapters in which they were found; thus, they may be difficult for readers to identify and internalize.

Meta-assessment rubrics. The meta-assessment rubric for Washington State University received the highest rating because it provided an example of assessing, making curricular or pedagogical modifications, and following-up to evaluate the results of those modifications. Gallaudet University and University of South Florida's meta-assessment rubrics also received high ratings because they contextualized use of results as changes to curricula and pedagogy, and noted the need to re-assess. Although space is limited on rubrics, these institutions have done well to use rubrics that explain use of results in greater detail, going above and beyond the most basic, first two criteria (*a direct reference to student learning or development*; and *a general mention of "use of results," "closing the loop," "improvement," etc.*).

Accreditation standards. None of the regional accreditation standards mentioned or described the need to re-assess or determine whether changes based on assessment results actually contributed to improvements in student learning. Furthermore, not one provided tangible examples, within the standards themselves, of what it means to use results. To be fair, some accreditors include additional information in more recent documentation. For example, in addition to information provided in the standards, the Southern Association of Colleges and Schools Commission on Colleges (2014) and Middle States Commission on Higher Education (2015) provide several guidelines and publications regarding accreditation, some of which detail student learning improvement. Yet, the standards themselves offer little guidance; more detail could be included without unduly lengthening them.

Until use of results is consistently communicated and understood, innovation in assessment practice and the ability to demonstrate improved student learning will likely stagnate.

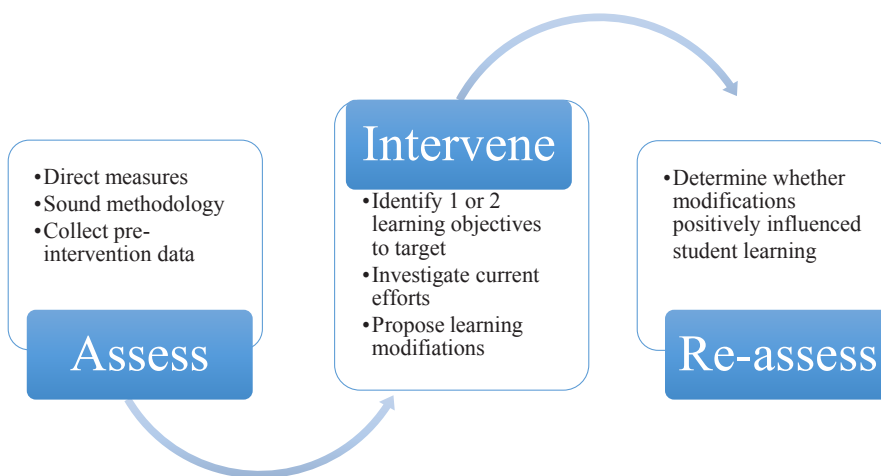


Figure 4. Exemplary definition of “use of results.”

After reviewing and rating various assessment resources in search of exemplary definitions of use of results, we found that all had shortcomings in reference to evidencing learning improvement.

Although some regional accreditors provide more detailed information about using results via additional documentation or publications, none of standards stated that re-assessment was a necessary part of using results to evidence learning improvement. Those involved with crafting regional accreditation standards were probably cognizant about the role of re-assessment. However, adding re-assessment as an accreditation requirement might have overwhelmed institutions that were still in the beginning or intermediate stages of their assessment practice.

Exemplary Definition of Use of Results to Improve Student Learning

After reviewing and rating various assessment resources in search of exemplary definitions of use of results, we found that all had shortcomings in reference to evidencing learning improvement. However, Fulcher et al. (2014) communicated a clear, consistent, and comprehensive definition of using results to improve student learning. A visual representation of this definition is provided in Figure 4. Unlike other resources, this resource is entirely dedicated to unpacking the term use of results. In this paper, Fulcher and colleagues defined use of assessment results to improve student learning as a program, department, or unit that:

1. **Assessed** using sound instruments that tightly align with programmatic student learning objectives and directly measure student learning;
2. **Intervened** by making evidenced-based curricular and/or pedagogical modifications at the program level;
3. **RE-assessed** using the same instrumentation; and
4. Found that student learning actually improved compared to pre-intervention assessment results.

According to Fulcher and colleagues' (2014) definition, using assessment results to improve student learning occurs "when a re-assessment suggests greater learning proficiency than did the initial assessment" (p. 5). Use of results is defined in terms of changes to curricula, pedagogy, and teaching. Moreover, the necessity to re-assess is explicitly described and incorporated into this definition of using assessment results for learning improvement. Most importantly, this definition of use of results includes a hypothetical example of an academic program that used assessment results to demonstrate improvement in students' oral communication skills. Thus, readers have a tangible example, concretizing and demystifying what use of results means.

Implications for Practice

Furthermore, we believe that clearly communicating use of results for programs at the onset can foster a sense of excitement that is rarely observed by focusing on logistical or procedural assessment details.

Often, assessment is performed in an effort to improve student learning. Unfortunately, assessment practitioners and program stakeholders rarely translate assessment results into action. It is even rarer for practitioners and stakeholders to re-assess students' learning to determine the effectiveness of actions taken in response to assessment findings (i.e., did changes or modifications actually improve student learning?). The current study investigated how use of results is communicated. We identified several areas of inconsistency and vagueness. That is, we demonstrated that authors, practitioners, and accrediting organizations use a variety of expressions and terms to define use of results. Unfortunately, few described closing the loop clearly, consistently, and comprehensively. Also, we noted that authors, practitioners, and accrediting organizations typically do not devote as much time and detail to describing use of results as they do on other aspects of the assessment cycle. To be fair, quality assessment must precede use of results; thus, it is understandable that some practitioners are focused on the brass tacks of assessment (i.e., the essential logistical and procedural details of the assessment process). However, it is imperative to clearly define and conceptualize use of results so that use can be realized after quality assessment is achieved.

Furthermore, we believe that clearly communicating use of results for programs at the onset can foster a sense of excitement that is rarely observed by focusing on logistical or procedural assessment details. By situating use of results to improve student learning at the forefront of assessment practice, higher education professionals will likely experience greater buy-in to the assessment process and foster greater efforts to use assessment results in a

meaningful way. We recognize that many institutions are already well along in their assessment work. We encourage assessment professionals at such institutions to re-focus their work by hosting professional development events (e.g., workshops, roundtables, presentations, etc.) for faculty members that present a clearer definition of use of results and concrete examples of what it means to use results for student learning improvement. Such events could highlight how previous work fits into this overall goal, while also helping faculty develop a more profound understanding of what it means to use assessment results to improve student learning. Until use of results is consistently communicated and understood, innovation in assessment practice and the ability to demonstrate improved student learning will likely stagnate.

We identified specific definitions that we hope will foster a better understanding of what it means to use assessment results for learning improvement, while also facilitating clearer, more consistent conversations among practitioners and stakeholders. As we communicate what use of results means, we hope that higher education will move one step closer to evidencing improved student learning.

As we communicate what use of results means, we hope that higher education will move one step closer to evidencing improved student learning.

AUTHOR'S NOTE: Since this article was submitted for publication, Megan Rodgers Good earned her Ph.D. in Assessment and Measurement from James Madison University, and accepted a full-time position as the Director of Academic Assessment at Auburn University. In addition, Elizabeth Hawk Sanchez earned her M.A. in Writing, Rhetoric, & Technical Communication from James Madison University.

References

- Allen, M. J. (2004). *Assessing academic programs in higher education*. Boston: Anker Publishing Company.
- Astin, A. W. (1985). *Achieving educational excellence: A critical assessment of priorities and practices in higher education*. San Francisco: Jossey-Bass.
- Atkinson, R. K., Derry, S. J., Renkl, A. & Wortham, D. (2000). Learning from examples: Instructional principals from the worked examples research. *Review of Educational Research*, 70(2), 181-214.
- Banta, T. W., Lund, J. P., Black, K. E., & Oblander, F. W. (1996). *Assessment in practice: Putting principles to work on college campuses*. San Francisco: Jossey-Bass.
- Barr, R. B., & Tagg, J. (1995). From teaching to learning: A new paradigm for undergraduate education. *Change: The Magazine of Higher Learning*, 27(6), 12-25.
- Blaich, C. F., & Wise, K. S. (2011, January). *From gathering to using assessment results: Lessons from the Wabash National Study* (NILOA Occasional Paper No. 8). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.
- Bourne, L. E., Goldstein, S., & Link, W. E. (1964). Concept learning as a function of availability of previously learned information. *Journal of Experimental Psychology*, 67(5), 439-448.
- Bresciani, M. J., Gardner, M. M., & Hickmott, J. (2009). *Demonstrating student success: A practical guide to outcomes-based assessment of learning and development in student affairs*. Sterling, VA: Stylus Publishing.
- Brown, S., & Knight, P. (1994). *Assessing learners in higher education*. Florence, KY: Routledge.
- Clarke, R. A. (2002). Learning outcomes: The bottom line. *Communication Education*, 51(4), 396-404.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46.
- Erwin, T. D. (1991). *Assessing student learning and development: A practical guide for college faculty and administrators*. San Francisco: Jossey-Bass Publishers.
- Ewell, P. T. (2009). *Assessment, accountability, and improvement: Revisiting the tension*. (NILOA Occasional Paper No. 1). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.
- Fulcher, K. H., Good, M. R., Coleman, C. M., & Smith, K. L. (2014, December). *A simple model for learning improvement: Weigh pig, feed pig, weigh pig*. (NILOA Occasional Paper No. 23). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.
- Fulcher, K. H., & Orem, C. D. (2010). Evolving from quantity to quality: A new yardstick for assessment. *Research and Practice in Assessment*, 5, 13-17.
- Fulcher, K. H., Swain, M. S., & Orem, C. D. (2012). Expectations for assessment reports: A descriptive analysis. *Assessment Update*, 24(1), 1-2, 14-16.
- Huba, M. E., & Freed, J. E. (2000). Learner centered assessment on college campuses: Shifting the focus from teaching to learning. *Community College Journal of Research and Practice*, 24(9), 759-766.
- Kuh (2011, January). Forward. In Blaich, C. F. & Wise, K. S. (2011, January). *From gathering to using assessment results: Lessons from the Wabash National Study* (NILOA Occasional Paper No. 8). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.
- Messick, S. J. (1999). *Assessment in higher education: Issues of access, quality, student development, and public policy*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Middaugh, M. F. (2009). *Planning and assessment in higher education: Demonstrating institutional effectiveness*. San Francisco: Jossey-Bass Publishers.
- Middle States Commission on Higher Education. (2011). *Characteristics of excellence in higher education: Requirements of affiliation and standards for accreditation* (12th ed.). Philadelphia, PA: Middle States Commission on Higher Education. Retrieved from <https://www.msche.org/publications/CHX-2011-WEB.pdf>
- Middle States Commission on Higher Education. (2015). *Publications: Guidelines for institutional improvement*. Retrieved from http://msche.org/publications_view.asp?idPublicationType=5&txtPublicationType=Guidelines+for+Institutional+Improvement

- New England Association of Schools and Colleges. (2011). *Standards for accreditation: Commission on institutions of higher education New England association of schools and colleges*. Bedford, MA: Commission on Institutions of Higher Education. Retrieved from http://cihe.neasc.org/downloads/Standards/Standards_for_Accreditation_FINAL_2011.pdf
- North Central Association of Colleges and Schools: The Higher Learning Commission. (2014). *The criteria for accreditation: Guiding values*. Retrieved from <http://ncahlc.org/Criteria-Eligibility-and-Candidacy/guiding-values-new-criteria-for-accreditation.html>
- Northwest Commission on Colleges and Universities. (2010). *Standards for accreditation*. Retrieved from <http://www.nwccu.org/Pubs%20Forms%20and%20Updates/Publications/Standards%20for%20Accreditation.pdf>
- Palomba, C. A., & Banta, T. W. (1999). *Assessment essentials: Planning, implementing, and improving assessment in higher education*. San Francisco: Jossey-Bass
- Schuh, J. H. (2009). *Assessment methods for student affairs*. San Francisco: Jossey-Bass Publishers.
- Schuh, J. H., & Upcraft, M. L. (2001). *Assessment practice in student affairs: An application manual*. San Francisco: Jossey-Bass Publishers.
- Southern Association of Colleges and Schools: Commission on Colleges. (2012). *The principles of accreditation: Foundations for quality enhancement* (5th ed.). Decatur, GA: Southern Association of Colleges and Schools: Commission on Colleges. Retrieved from <http://www.sacscoc.org/pdf/2012PrinciplesOfAccreditation.pdf>
- Southern Association of Colleges and Schools: Commission on Colleges. (2014). *Handbooks, manuals and guides*. Retrieved from <http://www.sacscoc.org/handbooks.asp>
- Suskie, L. (2010). *Assessing student learning: A common sense guide* (2nd ed.). San Francisco: Jossey-Bass Publishers.
- Walvoord, B. E. (2010). *Assessment clear and simple: A practical guide for institutions, departments, and general education* (2nd ed.). San Francisco: Jossey-Bass Publishers.
- Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Western Association of Schools and Colleges: Senior College and University Commission. (2013). *2013 handbook of accreditation*. Alameda, CA, Author.