Notes in Brief

Planning the intended use of data and identification of bottlenecks are two best practices that faculty and administrators can use when they conduct assessments for the combined purposes of accountability and improvement. Prior to data collection, they need to have a clear plan of how the results should offer worthwhile insights. Upon identification of bottlenecks to learning and efficient operation of units, faculty can develop appropriate action steps to address these trouble spots. Planning for the use of data should improve the assessment process itself. The process of identifying bottlenecks will mostly help to improve outcomes. This article gives examples of how both practices were used effectively for both student learning and operational outcomes. Using these two best practices led to enhanced decision-making ability that completed the assessment loop. The examples show improvement in student learning, increased retention rates, and more effective educational programs.

AUTHORS

Phyllis Blumberg, Ph.D.
University of the Sciences

Two Underused Best Practices for Improvement Focused Assessments

Currently, the primary purpose of assessment of student learning in higher education is to document what is occurring (Hutchings, Huber, & Ciccone, 2011; Ikenberry & Kuh, 2015). These assessment efforts are often done to comply with regional or specialty accreditation standards. Such assessments are and will remain essential. Since educators want their programs and institutions to become or remain accredited, they often document a very high percentage of outcomes as met or even exceeded expectations. Yet, another essential purpose of assessment is to make improvements. These accountability assessments may not lead to data that can be used for improvement. When most expected outcomes are met, there is no reason to try to improve or to make changes. Although necessary for improvement, faculty and administrators may be reluctant to conduct assessments that reveal a program's weaknesses. Faculty fear they will look bad when students do not meet their learning outcomes or programs do not reach their operational goals. Faculty actually look bad if they never try to improve (Massa & Kasimatis, 2017). Assessments become more valuable and useful when they combine both purposes of accountability and seeking improvement.

This article showcases two best practices that faculty and administrators can use when they are conducting assessments for the combined purposes of accountability and improvement: plan intended use of data and identify bottlenecks in student learning. While both practices can lead to improvements in student learning and more effective operations, they come from different sources. Planning the intended use of assessment data comes from the mainstream current assessment literature (Kuh et al., 2015). Identification of bottlenecks is an evidence-based practice that educational developers use to help faculty revise their courses (Pace & Middendorf, 2017).

CORRESPONDENCE

Email
p.blumbe@usciences.edu

Both practices can be used for the two common types of assessment: student learning and operational outcomes. However, these practices offer different improvement benefits. Planning for the use of data should improve the assessment process itself. The process of identifying bottlenecks will mostly help to improve outcomes. Changes to both the assessment process and the assessment outcomes are useful for both accountability and improvement assessment functions.

This article gives examples of how both practices were used effectively at the author's institution. This is a private, specialized, small university offering undergraduate and graduate degrees in the sciences and clinical professional degrees in the health sciences. The examples discussed come from recent assessment reports completed by directors of academic units.

Plan Intended Purpose of Assessment Data

Prior to data collection, faculty and administrators need to have a clear plan of how the results should offer insights about student learning or effective operation of units. Without such a plan, the data may not be relevant or may not be acted upon (Kuh et al., 2015). Although this seems like a common-sense idea it is not always used. Skipping how data will be used is not explicitly mentioned in the often referred to assessment cycle heuristic (Kinzie, Hutchings, & Jankowski, 2015; Suskie, 2009). When faculty just need to report on assessment data, as they might do in the accountability function of assessment, they may not have planned how the data will be used. In such cases they may just collect data that looks relevant to the program. However, assessments take on additional meaning once the faculty explicitly plan how they will use the data. Explicit plans for data use lead to more precise questions about how well the program is meeting its goals about student learning or efficient operations. Planning provides an anticipated idea for how data will be interpreted. Thus, it leads to better decision making (Kuh et al., 2015).

Assessments become more valuable and useful when they combine both purposes of accountability and seeking improvement.

The following two examples both use nationally normed, external exams as an appropriate measure of student learning. Prior to obtaining the data, when educators consider how they will use student performance data on these tests they are more likely to plan possible changes or action. However, the first example illustrates how an educational program initially collected data for accountability purposes without planning for its use. Once they identified the intended use of the data they were able to close the assessment loop both for accountability and improvement purposes. In contrast, the second example illustrates the planned use of data.

The biological sciences department requires that all graduating seniors take the ETS Major Field Test for biology as one of their major indictors of student learning. The requirement was to take the exam but the test had no impact on student grades or graduation. Students took this exam toward the end of their last semester and they did not take it seriously. While faculty were not pleased with the results they continued to require it because they felt they needed a valid, cumulative measure of student learning that was easy to administer. This is an example of conducting an assessment just to collect data for the sole purpose of accountability. Once the faculty asked what they would do with the student scores on this exam, they cared about student performance. This led them to make changes to try to improve performance. First, they moved the exam to an earlier semester. Students who performed significantly below the national average on the separate sections were asked to take another course relating to this content before graduating. Faculty assumed that this additional course should remediate these deficiencies. Upon further inspection of the results, the faculty found that many students, even some of their best students, were doing poorly on a few sections. This led the faculty to examine the alignment between their curriculum and the content on this national exam. They realized the exam was not a good indicator of mastery of the content emphasized in their major. Faculty are now considering using a different exam to measure cumulative student learning in their major. A possible operational outcome would be to identify or develop an appropriate and valid cumulative exam that aligns with their learning outcomes.

Prior to obtaining the data, when educators consider how they will use student performance data on these tests they are more likely to plan possible changes or action.

The faculty in the pharmacy program planned how they would use assessment data for both accountability and improvement purposes. For years the pharmacy program has been requiring students to take a test of mastery of pharmacy knowledge. The faculty use the results to gauge how well their students are doing in comparison to their national peers, as a stated student learning outcome. When repeated results indicated that pharmacy students toward the end of their first professional year of training were below the national norms, the faculty decided to change the curriculum to help the students master the required content earlier and better. At first, they made small changes in the scope and sequence of material. These changes did not lead to significant improvements on this exam. When small changes did not lead to improvement in student performance, faculty were motivated to totally revise their curriculum and how it is taught. The new curriculum fully integrates the basic pharmaceutical sciences with the clinical applications. Instead of the traditional lecture-based courses they will be using many more active learning techniques, especially team-based learning. The new curriculum is being implemented this year. The scores on this nationally-normed exam will be used as a major indicator of the success of the changed curriculum.

In addition to using nationally normed exams, many faculty use course-embedded assessments with intended purposes, as the following example illustrates. The general education program assessment plan explicitly states the intended use of the data, "The evidence will be used to make informed decisions about curriculum, pedagogy, assessment, and instructional resources". This program requires that all undergraduate students gain competency in six skills. Specific courses have been approved to teach and assess students on one or more of these skills. Students are motivated to take them seriously since the assessment activity is part of the course grade. As the assessment plan states, faculty-directed, course-embedded assessments were chosen because they are more likely to be used for curricular improvement.

Each skill is considered every three years and two skills are reported on annually. Faculty who teach these skill-approved courses report on cohorts of student performance using a course-specific, summative assessment instrument that measures this skill. These direct measures of student learning may take different forms but must include a four-point scoring scheme (1. not met; 2. approaching; 3. met; and 4. exceed expectations) for the student learning outcome(s). In 2017, faculty who taught courses that included ethics or oral communication reported on their assessments. Greater than 90% of the students were reported as meeting or exceeding expectations for both skills, with some faculty reporting extremely high levels of students exceeding expectations (e.g., 100%). While these high scores were fine for accountability purposes, considering improvement caused the general education committee to delve deeper into the meaning of these results. These committee members found that different instructors use different criteria for meeting these levels. As a result, they decided to hold faculty focus groups to talk about how the skills are assessed. These focus groups led to the development of clearer criteria for scoring student achievement; these criteria will be used to develop skill-specific rubrics which should be used by all instructors whose courses satisfy the skill. The goal is to develop assessment standards that are similar across different courses and instructors. Next, the general education committee will conduct professional development with faculty to calibrate the instructors' use of the rubrics. Such development should lead to a consistent application of rubric criteria across instructors and courses.

The goal is to develop assessment standards that are similar across different courses and instructors.

In addition to planning the use of data, results of assessments can help identify ways to improve programs. When the data indicate students are not doing as well as expected, faculty can try to find why these results were obtained. Identifying bottlenecks can be a useful method for determining where the problems are.

Identify Bottlenecks in Student Learning and Operational Effectiveness

Since the beginning of this century, faculty at Indiana University have been engaged in a process designed to increase learning (Pace & Middendorf, 2004). The first step in this process is to identify bottlenecks in student learning. Bottlenecks can be either cognitive or emotional. Cognitive bottlenecks relate to the difficulties students have with specific content. Cognitive bottlenecks create obstacles to student success and persistence in a discipline. Emotional bottlenecks relate to student anxieties or fears about the content. Math anxiety and religious beliefs that might promote resistance to the concept of evolution are good examples of emotional bottlenecks. Faculty in at least ten countries now are using this evidence-based process to identify ways to increase student learning (Pace & Middendorf, 2017). Although not referred to as a formal assessment method, identification of trouble spots is frequently used for continuous program improvement in higher education.

Instead of asking faculty to identify weaknesses in their programs, ask them to identify bottlenecks that impede student learning and success in educational programs. Bottlenecks can be found by inspecting where student cohorts struggle. This turns assessment into looking for ways to improve and does not the carry the negative connotation of weaknesses. When applied to programs, the identification of bottlenecks can be a practical tool that faculty and administrators can use in assessing programs.

The bottleneck concept has a long history in manufacturing improvement initiatives whereby managers identify where and why product creation is reduced. A similar process can be applied to educational programs. Faculty can identify bottlenecks by reviewing semester-to-semester retention rates, student grades, and comments. Upon identification of trouble spots in educational programs, people can develop appropriate action steps to release these bottlenecks. To address the recent concern about timely graduation rates (program or institution-wide bottleneck), higher education administrators have adopted various approaches to increase completion rates. Programs geared toward increasing retention of beginning students are common (Hart Research Associates, 2012).

Once the bottleneck has been identified faculty can make appropriate changes to the program that attempt to address these trouble spots. For example, a program might identify that many students do not master required mathematics skills. An analysis of the items that many students got wrong on these skills assessments would provide diagnostic information about which types of questions or content are difficult for students. Thus, the assessment data identifies specific concepts or skills that the students find especially hard to master. The faculty could explore if they could find a different way to teach these concepts or skills to make it less difficult for the students. After changing how they teach this content, a resulting student learning outcome might be to attain a 15% increase in the number of students who achieve mastery scores on those questions that relate to these identified mathematics skills across several courses that assess them. This program also could identify an operational outcome that increases student retention by 10% in the program. Identification of bottlenecks and making changes because of this knowledge may be less threatening for faculty than stating assessment in terms of vulnerabilities.

Identification of bottlenecks and making changes because of this knowledge may be less threatening for faculty than stating assessment in terms of vulnerabilities.

Like their colleagues across the country, the faculty at this university are concerned with retention in STEM (Felder & Brent, 2016), as it traditionally has been a barrier to students remaining in their intended major—whether that is in STEM or in health professions that require a good STEM foundation. At this specialized science and health science university all students must do well in STEM courses not only to stay in their major but also to remain at the university. For example, doing well in organic chemistry is required for not only chemistry majors but also pre-health professional students who aspire to become pharmacists and physicians. Faculty members have been employing reform efforts to teach using best practices in most of the STEM introductory courses. The reform efforts in the general chemistry course, described next, illustrate a sustained effort to identify and overcome bottlenecks in a gateway STEM course required for most of the students at this university. This assessment has been used for both accountability and improvement for years.

In general chemistry, prior to 2002, more than 30% of the first year students earned a D or F or withdrew from the course (DFW). The course involved weekly three hours of lecture, two hours of laboratory, and an optional one hour for recitation where students had the opportunity to ask questions and the professor demonstrated the solution to chemistry problems. The faculty reviewed the mistakes students made on the exams and found that a majority of students had the most trouble with higher-order questions where they had to apply concepts. Thus, problem-solving skills were the bottleneck. In 2002 the faculty changed the format of the recitation from a large class to mandatory smaller recitation sections where the students solved problems in small groups. This restructuring led to a 10% reduction in DFW grades (Mahalingam, Schaefer, & Morlino, 2008). The following year the students were required to do homework where they solve problems prior to coming to the recitation. While students came to class with their homework done, many still did not understand how to solve these problems. Upon questioning the students, they indicated they copied their answers from others, as the assignments were mandatory. The faculty hypothesized that implementing

While the dissertation is a major challenge for doctoral students, bottlenecks can occur at various stages of graduate education. Faculty should look at where attrition occurs to determine program-specific bottlenecks.

an online homework system should help overcome the bottleneck of understanding how to solve problems. Now the homework problem sets are more relevant to the exam questions, so students likely take the homework more seriously as meaningful preparation for exams, as opposed to busy work. This is a good example of aligning learning/practice activities with assessments. After experimenting with different online homework systems, faculty found that providing hints on how to solve problems throughout, and not just showing the steps of the problem, was the most helpful for student mastery of problem-solving skills (Mahalingam & Fasella, 2017).

The passive nature of the lecture classes also served as a barrier to problem-solving skill acquisition (Weimer, 2004). Since 2009, the faculty who teach this course have incorporated an audience response system to allow all students to answer questions throughout the lectures. The audience response system gives students immediate feedback that allows them to evaluate their understanding of the content and its application to problems.

The grading system also changed since 2002 when only exam grades and laboratory performance counted. Now, performance on homework and recitation problems counts toward the final grade. Therefore, final course grades are not valid comparisons. Instead, performance on exams is the appropriate pre- and post- educational intervention comparison. In addition, over time, the percent of application questions on each exam has increased. The percentage of students earning D or F grades on exams dropped from over 30% to 15% even as the exams got harder. This example shows how faculty can identify bottlenecks to student success and implement changes that result in significant increases in student learning and understanding of the content.

Retention and graduation rates are of even more concern in graduate education because nationally there is about a 50% attrition rate from PhD programs (Lovitts & Nelson, 2000). While the dissertation is a major challenge for doctoral students, bottlenecks can occur at various stages of graduate education. Faculty should look at where attrition occurs to determine program-specific bottlenecks.

The director of the master's degree program in biomedical writing developed an operational goal of a 75% graduation rate, which he measured for both accountability and improvement purposes. This is a reasonable graduation rate because this program attracts nontraditional students, most of whom are employed. Some students discover that the field of biomedical writing is not for them or decide that they want to pursue other careers. Recently, this program had a retention-to-graduation problem, as far fewer than 75% of the students graduated. Once this decreased graduation rate occurred, the director determined that most of the attrition occurred either during or at the end of the recommended first course. Between 25–50% of students were either dropping out during the course or not continuing to the next semester after taking this course. Therefore, the first course was this program's bottleneck. The program director decided to gather data about the course from the students who dropped out and from those who continued in the program. Other faculty and nonfaculty practitioners in the field also examined the syllabus. In addition, the director looked at student weaknesses in more advanced courses.

The data indicated that over the years, the instructor increased the required content and tried to raise the rigor of the course through several writing assignments which required accurate use of the American Medical Association (AMA) writing style. When a new, adjunct instructor began teaching this course she continued to implement these changes and even increased the expectations. The students perceived that the course was intended to weed out the less-qualified students, especially those who were not yet employed in biomedical writing. This perception is contrary to the philosophy and goals of the program which aims to give students the skills to be able to be employed in biomedical writing or to advance their biomedical writing careers. No courses are expected to eliminate less-experienced students. The conclusion of the faculty and the external reviewers was that the course was too ambitious for beginners. Those students who were already employed as biomedical writers were able to succeed with the assignments.

Because of this review the program director together with the instructor made a significant change in the content of the course. During this course the students are now taught how to write research reports using the industry's standard conventions, such as what goes into the introduction, methods, results, and discussion sections, and how to write using the AMA writing style—instead of assuming they knew how to do this. Some of the content was removed from this course and placed in a more advanced course. Since the implementation of these revisions, the dropout rate after taking this course fell to 5%.

These two examples show that identification of bottlenecks can be used for both student learning and operational outcomes. Once the bottlenecks are identified, the most critical step is to close the assessment loop by making changes to overcome the bottleneck. These changes can be made incrementally over a long period as the chemistry example illustrates or made quickly as was done with the biomedical writing example. In both cases, faculty were comfortable talking about assessments that showed previous students had struggled because they now fostered greater student success.

Discussion and Conclusion

The examples described here mirror the different types of recommendations that result from assessment (Massa & Kasimatis, 2017). As the examples show, assessments can lead to more than one type of recommendation. Course or curriculum revision occurred in pharmacy and biomedical writing. The faculty changed their pedagogy in chemistry and pharmacy. The general education assessment led to improved assessment of student learning, and a better alignment between the curriculum and the assessment tool. Repeatedly observing lower than expected student performance on exams can lead to different improvement action plans. Once the ETS Biology exam had a real purpose the faculty looked at the instrument itself. Since they were satisfied with their curriculum, they realized they needed an exam that aligned better with their learning outcomes. In the pharmacy example, the results suggested that the faculty needed to change their curriculum because the test was a valid measure of what the faculty expected the students to learn. By taking a deep dive into the data the faculty were able to close the assessment loop. The programs improved student learning and increased retention rates both in gateway undergraduate STEM courses and an introductory graduate course. Best of all, these improvements were made without needing many additional resources.

Student learning outcome assessment data also led to changes in operational goals, such as increased retention and graduation rates or curriculum revision.

These examples illustrate how faculty collect and study assessment data after planning the intended use of data or by identifying bottlenecks. Such data helped to determine whether student learning outcomes were met, which led to changes in what and how students were taught as well as how they were tested. Student learning outcome assessment data also led to changes in operational goals, such as increased retention and graduation rates or curriculum revision. The examples provide evidence for the framework used throughout this article: both common types of assessment (student learning outcome and operations) can support accountability and improvement purposes.

The two best practices discussed here—planning the intended use of data and the identification of bottlenecks—facilitate assessments for the dual purpose of accountability and improvement. Both practices encourage faculty to engage in meaningful student learning outcome and operational assessments. Perhaps the greatest benefit of these practices is that they are nonthreatening for those who use them. These practices do not make individual faculty members look bad or identify weaknesses of individual courses that could be held against individuals. The use of these practices reflects well on the people who use them because it shows they are trying to improve their programs and student learning. When provosts or deans actively promote the use of these best practices they are creating a supportive environment for meaningful assessments to occur.

References

Felder, R.M., & Brent, R. (2016). *Teaching and learning STEM: A practical guide*. San Francisco: Jossey-Bass.

Hart Research Associates. (2012). *The completion agenda: Post-secondary education leaders' perspectives on issues of strategies for increasing completion rate*.

Hutchings, P., Huber, M. T., & Ciccone, A. (2011). *The scholarship of teaching and learning reconsidered*. San Francisco: Jossey-Bass.

Ikenberry, S. O., & Kuh, G. D. (2015). From compliance to ownership: Why and how colleges and universities assess student learning. In G. Kuh et al. (Eds.), *Using evidence of student learning to improve student learning* (pp. 1–23). San Francisco: Jossey-Bass.

Kinzie, J., Hutchings, P., & Jankowski, N. A. (2015). Fostering greater use of assessment results. In G. D. Kuh et al. (Eds.), *Using evidence of student learning to improve higher education* (pp. 51–91). San Francisco, CA: Jossey-Bass.

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

Lovitts B.E., & Nelson, C. (2000). The hidden crisis in graduate education: Attrition from Ph.D. programs. *Academe, 86*(6), 44–50.

Mahalingam M., & Fasella E. (2017). Effective use of technology for asynchronous learning to elevate students' knowledge and problem solving ability. In S. P. Ferris & H. A. Wilder (Eds.), *Unplugging from the classroom* (pp. 149–158). Oxford, UK: Chandos Publishing (Elsevier).

Mahalingam, M., Schaefer, F., & Morlino, E. (2008). Promoting student learning through group problem solving in general chemistry recitations. *Journal of Chemical Education, 85*(11), 1577–1581.

Massa, L. J., & Kasimatis, M. (2017). *Meaningful and manageable program assessment*. Sterling, VA: Stylus.

Pace, D., & Middendorf, J. (2004). Decoding the discipline: A model for helping students learn disciplinary ways of thinking. *New Directions for Teaching and Learning, 98*(3), 1–12.

Pace, D., & Middendorf, J. (2017). Foreword to using the decoding the disceplines framework for learning across the dsiciplines. *New Directions for Teaching and Learning, 150*(3), 9–11. doi:10.1002/tl.20243

Suskie, L. (2009). *Assessing student learning* (2nd. ed.). San Francisco: Jossey-Bass.

Weimer, M. (2002). *Learner-centered teaching: Five key changes to practice*. San Francisco: Jossey-Bass.