



**Association for
Institutional Research**

PROFESSIONAL FILES | SUMMER 2017 VOLUME

Supporting quality data and decisions for higher education.

Letter from the Editor

Summer brings time to reflect and recharge. The Summer 2017 volume of AIR Professional Files presents four articles with intriguing ideas to consider as you plan for the next academic year.



Data governance is a pressing issue for many IR professionals, as sources of data proliferate and challenge our ability to control data integrity. In her article, *Institutional Data Quality and the Data Integrity Team*, McGuire synthesizes and interprets results from 172 respondents to an AIR-administered survey of postsecondary institutions on their data integrity efforts. She describes the current state of data governance and offers strategies to encourage institutional leaders to invest in data quality.

Those of us who work in assessment often take it for granted that assessment results will be used for learning improvement. Fulcher, Smith, Sanchez, and Sanders challenge this assumption by analyzing information from program assessment reports at their own institution. *Needle in a Haystack: Finding Learning Improvement in Assessment Reports* uncovers many possible reasons for the gap between obtaining evidence of student learning and using that evidence for improvement. The authors suggest ways to promote learning improvement initiatives, and share a handy rubric for evaluating assessment progress.

Institutional researchers are beset with requests to form peer groups, and it seems that no one is ever satisfied with the results. Two articles in this volume present very different methodologies for forming sets of comparison institutions. In her article, *A Case Study to Examine Three Peer Grouping Methodologies*, D'Allegro compares peer sets generated by different selection indices. She offers guidance for applying each index and encourages cautious interpretation of results. Rather than rummaging around for the perfect peer set, Chatman proposes creating a clone, or doppelganger university, one that is constructed from disaggregated components drawn from diverse data sources. In *Constructing a Peer Institution: A New Peer Methodology*, he walks us through the process of creating peers for faculty salaries, instructional costs, and faculty productivity. While the constructed peer approach has its challenges, the appeal of achieving a perfect fit peer is undeniable.

I hope your summer "reflection" inspires you to share your work with your IR colleagues through *AIR Professional Files*.

Sincerely,

A handwritten signature in cursive script that reads 'Sharron L. Ronco'.

Sharron L. Ronco

IN THIS ISSUE...

Article 140 <i>Author: Katherine A. McGuire</i> Institutional Data Quality and the Data Integrity Team	Page 1
Article 141 <i>Authors: Keston H. Fulcher, Kristen L. Smith, Elizabeth R. H. Sanchez and Courtney B. Sanders</i> Needle in a Haystack: Finding Learning Improvement in Assessment Reports	Page 19
Article 142 <i>Authors: Mary Lou D'Allegro</i> A Case Study to Examine Three Peer Grouping Methodologies	Page 35
Article 143 <i>Authors: Steve Chatman</i> Constructing a Peer Institution: A New Peer Methodology	Page 55

EDITORS

Sharron Ronco
Coordinating Editor
Marquette University

Leah Ewing Ross
Managing Editor
Association for Institutional Research

Lisa Gwaltney
Editorial Assistant
Association for Institutional Research

ISSN 2155-7535

NEEDLE IN A HAYSTACK

Finding Learning Improvement in Assessment Reports

Keston H. Fulcher
Kristen L. Smith
Elizabeth R. H. Sanchez
Courtney B. Sanders

About the Authors

Keston H. Fulcher is executive director of the Center for Assessment and Research Studies (CARS) at James Madison University, and associate professor of graduate psychology. Kristen L. Smith is doctoral graduate assistant for CARS. Elizabeth R. H. Sanchez is assistant for CARS. Courtney B. Sanders is doctoral graduate assistant for CARS.

Acknowledgments

Special acknowledgments and thanks to the following for their contributions: A. J. Good, Cathryn Richmond, and Alena Gordienko.

Keywords: Student learning improvement, use of results, student learning outcomes assessment, higher education assessment, assessment reporting

Abstract

Higher education insiders trumpet the use of results for improvement as the most important part of the assessment cycle. Yet, at the same

time, we acknowledge the rarity of improvement, especially at a program level. What are some reasons the most important phase of assessment occurs so infrequently? To seek answers, we investigated the “Use of Results” sections in 54 program-level assessment reports. In some respects, our findings were positive. On average, programs reported making approximately three curricular or pedagogical changes annually. A closer inspection, however, revealed concerns: (1) the curricular or pedagogical changes were not explicitly linked to learning outcomes, (2) programs rarely reported making changes that affect several classes, (3) many of the reported changes were unclear, (4) and few programs reassessed to determine if changes actually led to learning improvement. Our research concludes by providing suggestions for how programs can more effectively use results to inform changes, reassess students to determine if changes led to learning improvement, and report on improvement processes.

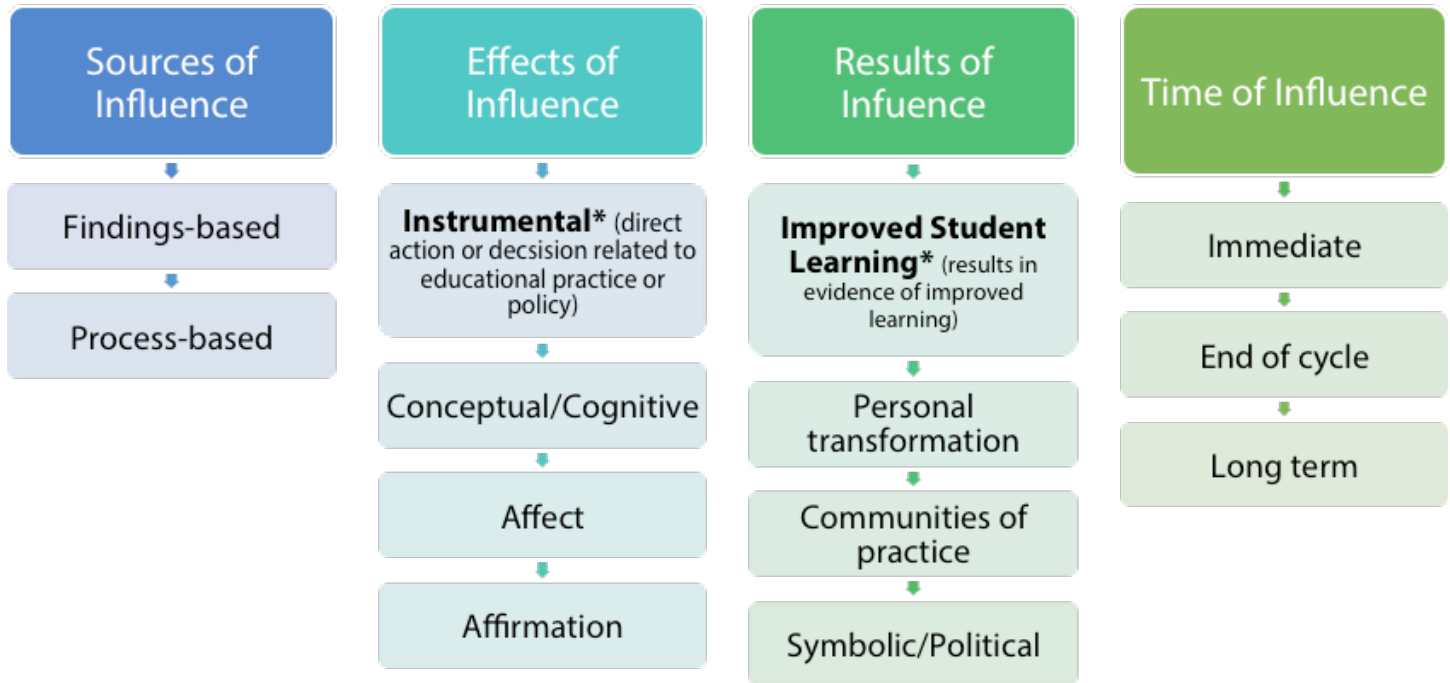
INTRODUCTION

For more than 30 years, higher education has refined assessment methodologies to meet accountability demands and demonstrate value

(Ewell, 2009). Yet, as Suskie (2010, para. 8) observed, “Today we seem to be devoting more time, money, thought, and effort to assessment than to helping faculty help students learn as effectively as possible.” Other researchers have come to a similar realization: although most institutions systematically collect assessment data, few use the data to improve student learning (Banta & Blaich, 2011; Blaich & Wise, 2011).

Why aren’t assessment results used for learning improvement? There are several theories: It could be that institutions incorrectly assume that using results for improvement can emerge from only interesting research findings and well-crafted reports (Blaich & Wise, 2011). It could also be that inconsistent and vague communication surrounding the use of results for improvement confuses programs (Smith, Good, Sanchez, & Fulcher, 2015). Furthermore, accreditation requirements, rather than intrinsic interests, might be the main driver of assessment practices (Kuh & Ikenberry, 2009). Indeed, a myopic focus on assessment activities (e.g., identifying outcomes and gathering data) unintentionally neglects using results for student learning

Figure 1. Depiction of Current Study Within Jonson et al.'s (2014) Heuristic Model of Influence



improvement (Kinzie, Hutchings, & Jankowski, 2015).

Not using results to inform curricular and pedagogical changes remains a serious problem for higher education. To investigate the issue, we analyzed "Use of Results" sections in 54 assessment reports. While the current study emphasizes learning outcomes assessment at the academic degree program (e.g., bachelor's degree in biology), many concerns and findings likely generalize to other assessment and institutional effectiveness initiatives. Indeed, the inability to use results to make changes that promote improvement is an institutional concern.

Conceptualizing Use of Assessment Results


We are not the first assessment practitioners to examine why using results to improve student learning remains uncommon. For example, Jonson, Guetterman, and Thompson (2014) believed that higher education could benefit from a new, broader definition of use of results.

Instead of focusing on curricular and pedagogical changes intended to improve student learning, Jonson et al. (2014) created a model to describe various ways that discussing results can positively influence the culture of a university (Figure 1). For example, using assessment results for discussion

can support taking direct action on educational practice or policy or changing people's ways of thinking about learning and assessment. Results for discussion can also alter people's emotions or attitudes regarding assessment practice and affirm the efficacy of an existing practice.

Jonson and colleagues (2014) further explained that each of the aforementioned influences could lead to the following outcomes:

- Evidence of improved student learning
- Transformation of stakeholders
- Building new communities of practice
- Generating support for policies and practice



The Jonson et al. (2014) framework sparks important conversations about how to define and measure using results for improvement, but we believe that a narrower, student-focused approach to using results would be of greater benefit to higher education. We define using assessment results for improvement as collecting and analyzing student learning data to support taking direct actions

related to educational practice (i.e., making changes to curriculum and/or pedagogy) that lead to evidence of improved student learning (i.e., students' assessment scores show improvement after experiencing modified curriculum or pedagogy).

Adopting the more-narrow definition of use of results, one that centers on student learning improvement, allows

us to keep in mind the overall intention of assessment and higher education. The current study is situated in the more-narrow definition, which might explain why we found so few examples of using assessment results in the 54 reports we examined.

Box 1. Hypothetical Example: 1980s Pop Culture Degree Program

At the conclusion of the 1980s Pop Culture degree program, students must be able to properly cite and reference a variety of sources in a research paper. In 2014–2015 the program used a rubric to evaluate all students' final research papers. Rubric scores revealed that students were not successful at citing or referencing sources. During a departmental discussion, program faculty confirmed that many students struggle to properly cite and reference sources.

After agreeing that the learning outcome of properly citing sources was both relevant and unmet, faculty agreed on curricular and pedagogical changes to address the issue. Before implementing new changes, faculty consulted with other instructors on campus and gathered information regarding what assignments could be effective at teaching such a specific skill set. Changes to the core courses of the 1980s Pop Culture program began in the fall of 2015. Specifically, the instructors of the two classes where writing is heavily emphasized—PCUL401 (1980s Politics and Culture) and PCUL404 (The 1980s and Today)—did the following:

1. Participated in a faculty development workshop during which the instructors found and agreed on examples of students' citing and referencing sources in their papers. Some examples were developing papers and others were advanced papers.
2. Shared the results of the past writing assessment with students, emphasizing that citing and referencing sources is a concern.
3. Provided modified examples of a developing and advanced paper to illustrate program expectations.
4. Created more in-class assignments to measure student progress, and encouraged students to rely on their own skills, instead of on online citation software, to create references.
5. Used the writing rubric to evaluate students' essays throughout the semester instead of using the rubric solely for the final research paper.

Results from curricular and pedagogical changes suggested that students' ability to cite and reference sources, as measured by the writing rubric, improved over time. Specifically, seniors' scores on the citing and sourcing element increased from 2.6 (between developing and competent) in 2015, the year before the curricular and pedagogical changes were implemented, to 3.2 in 2016 and 3.4 (between competent and advanced) in 2017, the years after the changes were implemented.

Understanding the Use of Results for Learning Improvement in Assessment Reports

Every year, academic programs at universities nationwide complete assessment reports that include a “Use of Results” section (Fulcher, Swain, & Orem, 2012). The current study examined the contents of these sections. More specifically, we investigated if changes to curricula or pedagogies were made based on assessment results and whether previous changes led to student learning improvement.

To evaluate the degree to which assessment reports conveyed using results for improvement, we first identified several ideal features of the “Use of Results” section in the assessment reports:

- Changes to curricula and pedagogies are made and reported.
- Changes to curricula and pedagogies are matched with an intended student learning outcome (i.e., what students should know, think, or be able to do).
- Changes to curricula and pedagogies are presented with a clear rationale (e.g., assessment data support changes).
- Reassessments demonstrate learning improvement (i.e., changes are at the program level and are effective).

To make the ideal assessment report more concrete, we provide an example from a hypothetical example: the 1980s Pop Culture degree program (Box 1).

The Current Research Study

Understanding how assessment reports could ideally connect assessment results to learning improvement efforts via curricular and pedagogical changes is important. We provided one simple example of a hypothetical program in an effort to clarify what the “Use of Results” section could, and should, include.

The current study focused on real programs attempting to use assessment results. We reviewed and qualitatively rated 54 program reports, comparing their features to our ideal assessment report. In doing so, we addressed five research questions (RQs).

Research Questions

RQ 1. How extensive in magnitude are the reported changes to curricula and pedagogies?

As we have explored, institutions and academic degree programs can use assessment results in different ways. Some use the results to inform changes to assessment instrumentation, while others use results to influence curricular and pedagogical changes. For those who used assessment results to change program curricula or pedagogies, we wanted to gauge the magnitude of the changes made, as described in assessment reports. That is, we wanted to see if the change was a course-level or a program-level change. If more students experience new curricula and pedagogy, we would expect to see more learning improvement at the program level.

We defined and evaluated magnitude of change in terms of minor, moderate, major, or extensive changes. An example of a change coded as minor in magnitude could include a new or modified course assignment based on previous assessment results. A change of moderate magnitude could be a new or modified unit or segment of the course curriculum. Major changes could entail a complete redesign of an entire course. Finally, extensive changes necessitate a restructuring of the curriculum or pedagogical approaches that involved several courses within a given academic program.

Again, we thought that perhaps programmatic changes of greater magnitude would be more likely to yield improved student learning. If faculty members are reporting that they only implemented changes of minor to moderate magnitude, this could help explain why no demonstrable student learning improvement exists. That is, using results to initiate only a minor or moderate change to curriculum, such as changing an assignment or unit in one course, might not be enough to move the needle at the program level.

RQ 2. To what extent are curricular and pedagogical changes linked to student learning outcomes?

To successfully improve student learning in a demonstrable way, faculty should focus assessment, pedagogical, and curricular efforts around specific student learning outcome(s) (Fulcher, Good, Coleman, & Smith, 2014). Once the learning outcome is identified, it should be

clear how curricular and pedagogical modifications would enhance students' skills, knowledge, or abilities.

We defined and evaluated the match between changes and student learning outcomes, differentiating among four levels of connection in the "Use of Results" sections we evaluated:

1. It might be unclear how the change is linked to student learning.
2. It might be that the change is linked to student learning in general, but not directly to a specific student learning outcome of the program.
3. It might be that the change is linked to a specific, program learning outcome and yet lack specificity about why or how the change aligns with that particular learning outcome.
4. It might be that the change is clearly linked to a specific learning outcome in such a way that improvement seems likely.

Demonstrable program-level learning improvement can be achieved only through changes that match student learning outcomes. In other words, if we cannot determine what students should know, think, or be able to do as a result of the programmatic changes, how will we know if the changes were successful at improving student learning? Programs that can align changes with student learning outcomes in a clear and logical way should have greater success evidencing improvement.

RQ 3. What is the rationale behind curricular and pedagogical changes?

Often, there are numerous reasons that programs decide to implement changes to curricula or pedagogies; it is important to explain the rationale for making specific pedagogical and/or curricular changes (Fulcher et al., 2014). Ideally, the rationale provided in assessment reports is not only explicit, but also originates from different sources (e.g., direct assessment measures, accreditation recommendations, etc.). It is plausible that when changes lack robust supporting rationale, they are less likely to culminate in demonstrable student learning improvement. A lack of understanding or articulation of the rationale for curricular and pedagogical changes might contribute to why minimal learning improvements are found in assessment reports.

We defined and evaluated the rationale for curricular and pedagogical changes provided in assessment reports based on explicitness and type. For explicitness, we coded the report rationales as either stated, but not explained or stated with an explicit rationale. For type, we determined whether the source that contributed to the rationale was a direct measure, an indirect measure, anecdotal (e.g., conversations), accreditation or annual program review recommendations, or realignment of instruction with changes in programmatic learning objectives.

RQ 4. What is the typical stage of implementation for curricular and pedagogical changes?

Curricular and pedagogical changes take time to implement. For instance, Fulcher and colleagues (2014) suggested that it could take 3 to 5 years to make program-level adjustments and subsequently use assessment results to demonstrate improved student learning. In addition to time, change requires planning and foresight. In order to coordinate change efforts, programs should create an improvement timeline. Timelines articulate when baseline assessment data will be collected, when pedagogical or curricular changes will be implemented, and when students will be reassessed to determine whether their learning actually improved (Fulcher et al., 2014).

It could be the case that programs conceptualize processes of curricular and pedagogical changes 1 year at a time—correlative of the assessment reporting cycle. We encourage programs to look beyond an annual cycle. Creating a 3- or 5-year plan and timeline might help motivate programs to use assessment results, make changes, and reassess students to demonstrate improved learning.

For the current study, we defined and evaluated the stage of implementation of change in terms of five criteria. Change efforts could be in one of the following five stages:

1. Planning (a program is currently planning changes);
2. In process (a program is currently implementing changes; some

changes but not all have been made);

3. Completed but have not yet reassessed;
4. for efficacy (or effectiveness reassessed) but no demonstrable improvement evidenced; or
5. Completed and checked for efficacy (or effectiveness reassessed) and demonstrable improvement evidenced.

RQ 5. To what degree are programs able to close the assessment loop by using results to inform changes and subsequently demonstrate improved student learning?

The promise of quality assessment practice is to enhance learning for students and improve higher education. That is, if programs are typically unable to close the assessment loop by using results to inform changes and demonstrate learning improvement, then assessment practice is falling short of its promise.

We addressed RQ 5 via the fifth stage of implementation criteria discussed previously for RQ 4. More specifically, change efforts coded as being at Stage 5 of implementation represented instances of closing the assessment loop (i.e., change efforts coded as “Stage 5: Completed and checked for efficacy (or effectiveness reassessed) and demonstrable improvement evidenced” were used to address RQ 5).

METHOD

Our home institution is a mid-sized, 4-year, public university in Virginia. The State Council of Higher Education

for Virginia (SCHEV) and our regional accreditor (Commission on Colleges of the Southern Association of Colleges and Schools, or SACSCOC) require colleges and universities to assess student learning. In compliance with their respective policies and guidelines, all academic degree-granting programs at our institution submit annual assessment reports for student learning outcomes. Each year graduate students, faculty members, and assessment specialists evaluate these assessment reports. Through feedback and consultation, several programs at our institution have demonstrated better assessment processes (Rodgers, Grays, Fulcher, & Jurich, 2013).

For this study, we examined all 54 exemplary assessment reports collected from the fall 2012–2013 reporting cycle. Fifty-four represents approximately half of our academic degree and certificate programs. Exemplary assessment reports received a score of 3.4 or higher out of 4, on a meta-assessment rubric (see Appendix A) (Fulcher & Bashkov, 2012; Fulcher & Orem, 2010). The 3.4 standard was set in 2011 by trained faculty using a modified Angoff procedure.

Our review included only exemplary assessment reports for practical reasons; we hypothesized that academic programs with established, high-quality assessment processes might be best poised to use assessment results to influence pedagogical and curricular changes (and subsequently demonstrate learning improvement). They also might be better equipped to reassess

students’ learning to determine if the implemented changes actually promoted learning improvement. Furthermore, programs in nascent stages of assessment (not close to exemplary) are likely focused on setting up assessment infrastructure. Such programs are typically establishing learning objectives, creating curriculum maps, and selecting assessment instruments. These programs, therefore, are less likely to have collected data and synthesized them into actionable results. Of course, use of results is a moot point to those programs that have not collected data. In essence, by focusing on exemplary reports we could rule out undeveloped assessment practices as an explanation for not using results to improve student learning. Within each exemplary assessment report, we identified specific descriptions of using results for improvement and then used an online Qualtrics survey to code each of the identified descriptions.

Procedures for Identifying and Coding Descriptions of Results

To locate specific descriptions of using results for improvement, a graduate student familiar with the meta-assessment rubric (see Appendix A) and the layout of assessment reports reviewed electronic copies of all 2012–2013 assessment reports in alphabetical order according to program name. The graduate student first read Section 6A, “Program Modification and Improvement Regarding Student Learning and Development,” of the assessment

report; this section asks program assessment practitioners to describe use of assessment results for student learning improvement. If there were no examples or evidence of use of results to improve student learning in Section 6A, the graduate student reviewed other sections in the reports. If the graduate student initially found no evidence of use of results, she set the report aside. Later, she rereviewed the report, reducing the chance of an overlooked example.

For each assessment report, the graduate student identified up to four examples that described use of results by electronically highlighting sections of the report in yellow. Note, of the 54 exemplary assessment reports, there were only two that had more than four examples. After the initial review and electronic highlighting, the graduate student randomized the order of the assessment reports and rereviewed them, converting the yellow highlighting to highlighting in red, yellow, green, or blue (i.e., example one was highlighted in red, example two in yellow, etc.).

Once the graduate student had reviewed all 2012–2013 exemplary assessment reports and highlighted all identified descriptions of use of results for improvement, three authors of this paper—along with three other graduate students—Independently evaluated and coded the using-results descriptions via an online Qualtrics survey. Specifically, raters reviewed all highlighted descriptions—each representing an individual “use of results”—in their assigned assessment reports. The raters

evaluated the following aspects of the descriptions:

- Magnitude of change, defined by extent or magnitude of changes made to pedagogy, curricula, and so on (minor: changes to a small class assignment in one class; moderate: change to a unit within a class; major: major overhaul of a class; extensive: numerous changes that affect several classes)
- Extent to which faculty linked change to student learning objectives
- Rationale for needing change
- Reported stage of change implementation

The six raters were paired into three groups of two; each group was assigned a subset of the 54 exemplary assessment reports. Groups 1, 2, and 3 evaluated a total of 20, 21, and 13 different assessment reports, respectively. First, each rater

independently coded the highlighted sections in every assigned assessment report, then each rater pair adjudicated to reach exact agreement on all coded sections. For instance, Raters 1 and 2 were paired and assigned 20 assessment reports to review; one of those assessment reports was from the Assessment & Measurement Ph.D. program. Each rater independently reviewed every highlighted description of using results within the Assessment & Measurement program assessment report. Then, using a Qualtrics survey, each rater coded the highlighted descriptions. Finally, they reviewed each other’s ratings and adjudicated until they agreed on all ratings for the Assessment & Measurement report. Each rater pair repeated this process for every assigned assessment report.

RESULTS

Across the 54 assessment reports, we identified and evaluated 162 different descriptions of using

Figure 2. Distribution of Magnitude of Curricular and/or Pedagogical Changes Across All Coded Program Assessment Reports

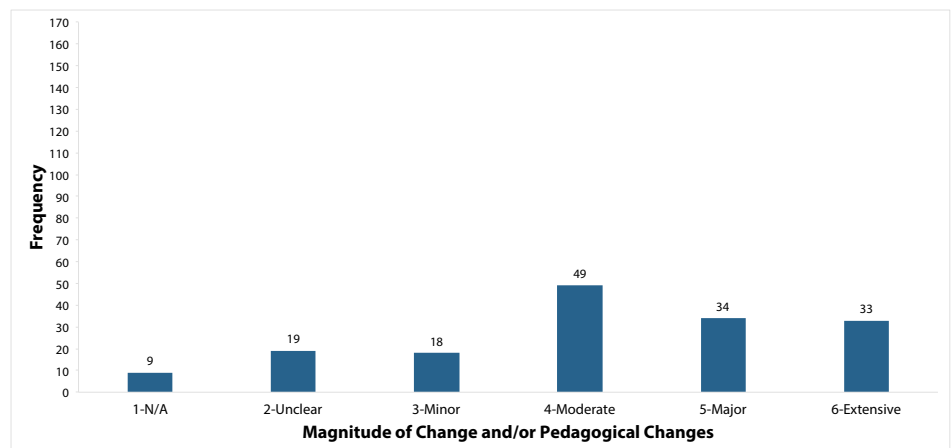
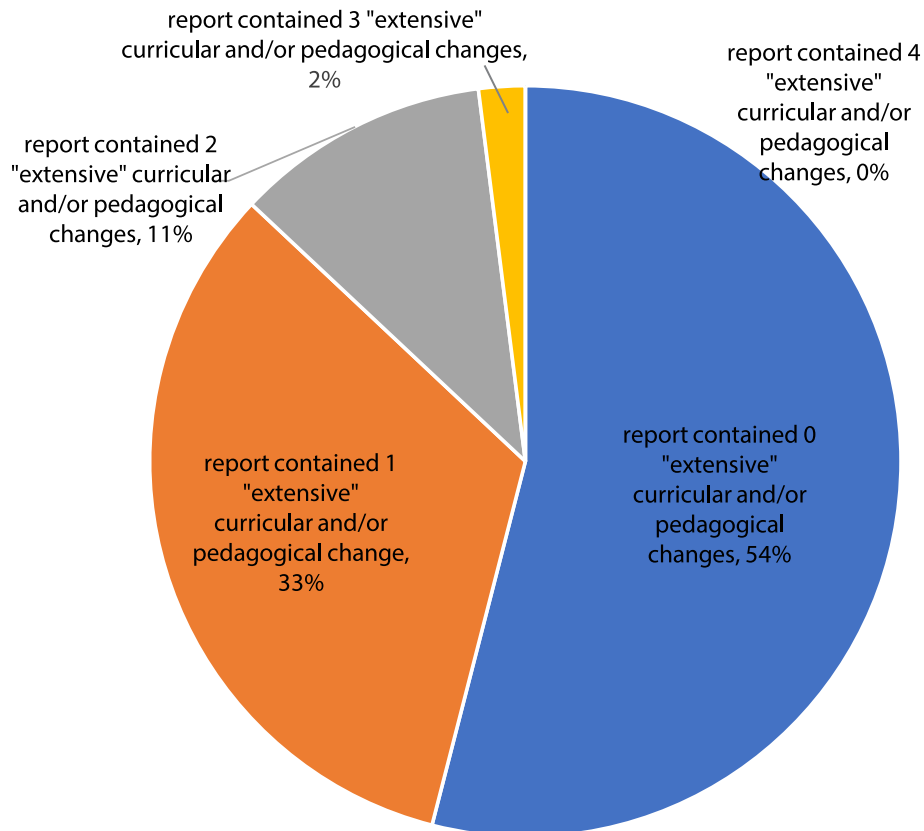


Figure 3. Percent of Program Assessment Reports That Contained Different Numbers of Curricular and/or Pedagogical Changes Coded as Being Extensive in Magnitude (e.g., Percent of Program Reports That Had Either 0, 1, 2, 3, or 4 Extensive Changes)



assessment results to make curricular or pedagogical changes. On average, we identified three descriptions per assessment report ($M = 3.00$, $SD = 1.13$). Clearly, reporting assessment data spurs talk of change. Nevertheless, only 8% of programs (among the 54 reports) could show that their pedagogical and curricular changes led to better learning outcomes. The following research questions (RQs) explore why so little learning improvement was reported despite so many changes within programs.

RQ 1. How extensive in magnitude are the reported changes to curricula and pedagogies?

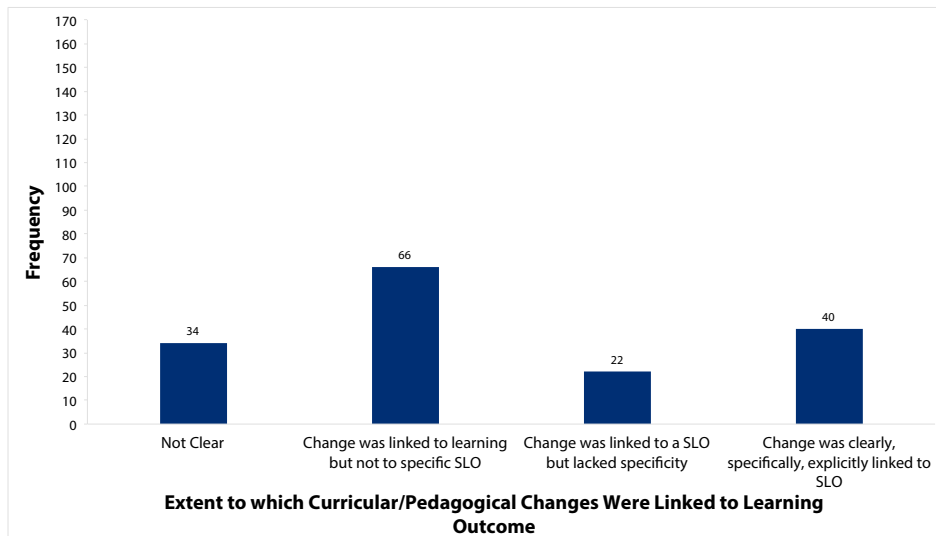
Recall, researchers rated the magnitude of curricular and pedagogical changes described in programmatic assessment reports based on the reported magnitude of changes made to courses, curricula, pedagogies, and so on. For instance, a change that involved only minimal adjustments to one assignment in one course would be rated as minor. Such an adjustment would not be expected to have a demonstrable, positive effect

on student learning, at the program or departmental level. Comparatively, a change that involved extensive modifications that affected multiple courses within the program or department would be expected to have a more demonstrable influence on program- or department-level student learning.

The magnitude of curricular and/or pedagogical changes was slightly negatively skewed (see Figure 2). In other words, the majority of the identified changes were coded as either moderate (a coded score of 4), major (a coded score of 5), or extensive (a coded score of 6) in magnitude. On average, the described changes were coded as moderate ($M = 4.10$, $SD = 1.45$). Nearly 20 of the identified changes were coded as unclear because, although faculty described a change, they did not provide enough information about the change to accurately identify its magnitude. For assessment reports in which faculty said they made a change, but then included no description of the change whatsoever, researchers applied the code "N/A."

Within each program assessment report, 54% had zero curricular and/or pedagogical changes coded as extensive in magnitude (see Figure 3). About 33% of the 54 assessment reports had one change coded as extensive in magnitude, 11% had two such extensive changes, and 2% had three. In addition, none of the 54 assessment reports contained four changes coded as extensive in magnitude. In essence, nearly half (46% of programs) reported the type of

Figure 4. Frequency of Identified Curricular and/or Pedagogical Changes That Were Linked or Aligned to Student Learning Outcomes



extensive pedagogical and curricular changes most often associated with learning improvement. However, these extensive changes equated to fewer examples of learning improvement than one might expect: only 8%. The results for RQs 2 to 4 provide more explanation to why these extensive changes led to so few examples of evidenced improvements.

RQ 2. To what extent are curricular and pedagogical changes linked to student learning outcomes?

Typically, curricular and pedagogical changes were linked to student learning generally (a coded score of 2), but were not matched to a specific, program-level student learning outcome ($M = 2.40, SD = 1.08$). As Figure 4 shows, approximately 34 out of the 162 identified curricular or pedagogical changes, or 21%, did not include enough details for raters to

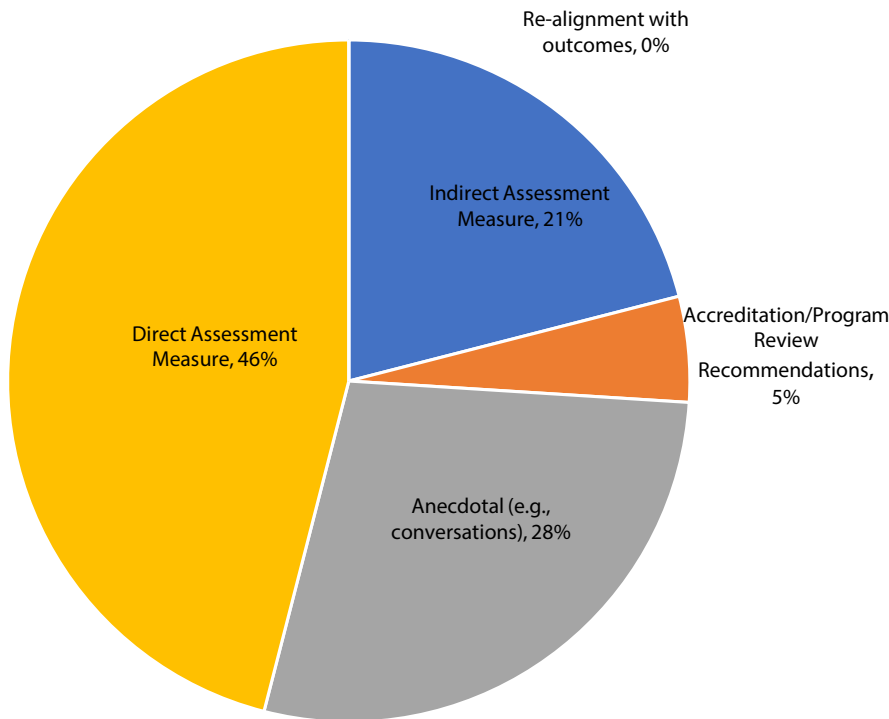
evaluate the alignment between the change and the program’s student learning outcome(s). The lack of explicit alignment between changes and student learning outcomes might be contributing to the issue at hand: insufficient use of assessment results to evidence improved student learning.

For many, the link between curricular and pedagogical changes and specific student learning objectives might be implicit. However, documenting the use of assessment results to influence pedagogical or curricular changes that lead to improved student learning requires explicit connections between implemented changes and student learning outcomes. It seems that assessment practitioners and support services need to better conceptualize and articulate the importance of matching changes to student learning outcomes.

RQ 3. What is the rationale behind curricular and pedagogical changes?

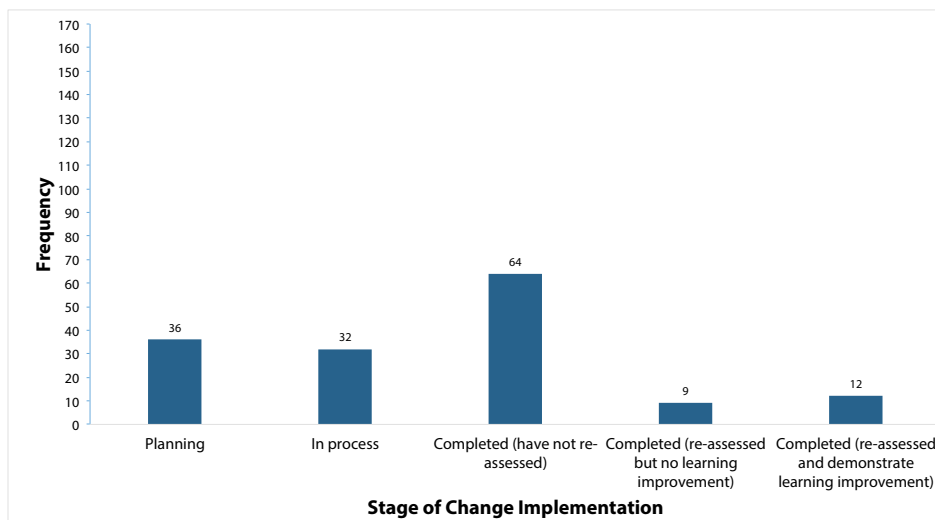
About 80% of the identified descriptions of curricular or pedagogical changes provided a rationale that conveyed the need for change. But just over 50% of the descriptions of curricular or pedagogical changes provided a rationale and mentioned the source that supported the rationale (i.e., direct assessment measures, accreditation or program review recommendations, etc.). In addition, about 19% of the identified descriptions of curricular or pedagogical changes provided no rationale. The most frequently provided rationale behind the described curricular or pedagogical changes was data from direct assessment measures. In contrast, few cited accreditation/program review as a rationale for a given change; none mentioned curriculum realignment. Of the program assessment reports that provided a source explaining their intended curricular and/or pedagogical change, Figure 5 displays the percent of reports that cited various sources of rationales for changes. Perhaps programs recognize the results of direct assessment measures, instead of feedback from accreditation/program reviews, as potential sources for change. In addition, some did not include any rationale to support changes to pedagogies and/or curricula. Perhaps the importance of understanding and describing the driving forces behind program-level changes is not recognized. Or, what might be a supportive rationale is not included because the report writer(s) believed the rationale was implied.

Figure 5. Of the Program Assessment Reports That Provided a Rationale and Source Explaining Their Intended Curricular and/or Pedagogical Change, Percent of Reports That Cited Various Sources of Rationales for Changes




Note, the meta-assessment rubric used at our institution in 2012–2013, the year of these reports (see Appendix A), does not require an explicit rationale to support curricular or pedagogical changes. Nonetheless, explicitly describing the rationale underlying change is an essential part of using results to demonstrably improve student learning (Fulcher et al., 2014). Given assessment measures were the most frequently cited rationale for curricular and pedagogical changes, intrinsic buy-in for change might be nonexistent. Alternatively, curricular and pedagogical changes that lack adequate rationale might not be well aligned with students’ learning needs, program resources, faculty sentiments, or administrative agendas.

Figure 6. Distribution of Stage of Change Implementation Ratings Across All Coded Program Assessment Reports



RQ 4. What is the typical stage of implementation for curricular and pedagogical changes?

Encouragingly, about 56% (85 out of 153) of the described curricular and pedagogical changes were complete. Yet, only 14% (21 out of 153) of all described curricular and pedagogical changes included follow-up reassessments (see Figure 6). Again, in 2012–2013 the crucial reassessment phase had not been explicitly stated in our institutional assessment cycle nor in our meta-assessment rubric (see Appendix A). Therefore, programs might not have been aware of the importance of reassessing. Alternatively, many might mistakenly believe that assessment work is done as soon as data are used for curricular and pedagogical change.



As Fulcher and colleagues (2014) explain, assessment practitioners, faculty members, and other stakeholders often confuse program changes with program improvements. A change is only an improvement when, upon reassessment, students demonstrate greater proficiency. Essentially, merely implementing curricular or pedagogical changes does not provide demonstrable proof of improved student learning, just as a pig never fattened because it was weighed. Assessment practitioners can do a better job of articulating and promoting the use of assessment results for improved student learning.

RQ 5. To what degree are programs able to close the assessment loop by using results to inform changes and subsequently demonstrate improved student learning?

As foreshadowed at the beginning of this section, only 8% of the evaluated curricular and pedagogical changes were implemented, reassessed, and demonstrated improved student learning. Our interpretation of this finding is that either programs are not closing the loop or our university programs do not know how to articulate such a process in an assessment report. Little integration of assessment processes with pedagogy and curricular design suggests a lack of clarity about learning improvement within our institution.

CONCLUSION

Even after more than 25 years of assessment practice at our university, finding evidence of student learning

improvement in assessment reports is akin to finding a needle in a haystack. To understand more about this most important phase of the assessment cycle, we qualitatively reviewed and coded 54 exemplary assessment reports from academic programs across our campus. In these assessment reports, writers described changes to course scaffolding, use of different classroom pedagogies, course redesigns, and so on. Furthermore, the curricular and pedagogical changes described were typically coded as being moderate in magnitude and were primarily driven by data from direct assessment measures.

However, under scrutiny, the thread from the “Use of Results” section to demonstrable student learning was typically thin and loose. Few programs could demonstrate the positive impacts of the curricular and pedagogical changes they made. Based on descriptions in the assessment reports, programs rarely conducted follow-up reassessment research to determine whether curricular and pedagogical changes had a demonstrable impact on student learning outcomes. Perhaps this finding can help explain why use of assessment results has not contributed enough to improving student learning outcomes in higher education (Kuh, Jankowski, Ikenberry, & Kinzie, 2014).

The inability to empirically demonstrate improved student learning was not for lack of earnest efforts to improve. That is, some programs conceptualized curricular and pedagogical changes, provided some rationale to support these changes, and implemented the

changes in their entirety. Yet, many of the program assessment reports lacked one or more critical elements, including

- Major or extensive pedagogical changes (i.e., changes at the program level);
- Tenable links between curricular and pedagogical changes and student learning outcomes;
- Convincing rationales to support curricular and pedagogical changes; and
- Adequate reassessment processes that can determine whether changes actually improved student learning.

Assessment Practitioners’ Role in Bridging the Gap between Using Results and Demonstrating Student Learning Improvement

In general, higher education stakeholders have not successfully evidenced systematic improvements in student learning at the academic program level. While making some progress, our institution certainly struggles. From a policy perspective, being a good shepherd of resources suggests that institutions are making earnest efforts to improve. Academe’s lack of demonstrating such improvement definitely contributes to the “Is college worth it?” conversation (Taylor et al., 2011).

To answer questions of worth and demonstrate the value of a college education, assessment results need to influence pedagogical and curricular changes at a program level. Ultimately, explicit gains in student learning

should be clearly articulated via assessment reports, presentations, and other channels of dissemination. Assessment practitioners must do more to communicate the importance of student learning improvement initiatives.

Findings from the current study reflect Blaich and Wise's (2011) observation that excellent assessment—by itself—does not lead to learning improvement. In addition, our results suggested that practitioners could increase student learning improvement by helping programs

1. Develop and implement more widespread and multiyear curricular and pedagogical changes;
2. Situate improvement efforts within student learning outcomes;
3. Understand the important role of reassessment; and
4. Use a framework or step-by-step example to more effectively report and explain crucial information.

As Fulcher and Bashkov (2012) explain, we should not be surprised that assessment reports lack adequate descriptions of using results to demonstrate improved student learning. At our institution, we did not offer enough guidance with respect to how to report our learning improvement efforts. In addition, we realized that we have no assessment staff trained in pedagogy, curriculum, course redesign, course scaffolding, or organizational change.

Lacking a holistic expertise within our own assessment office led us to engage in more-intentional partnership with our campus faculty development center. Doing so allows us to better serve faculty members as they create and implement curricular and pedagogical changes, and then reassess students' learning. For instance, the faculty development experts assist programs as they articulate student learning outcomes and align them with program theory.

We hope that the recommendations from the current study can assist institutions in better conceptualizing, articulating, implementing, reporting, and disseminating learning improvement success stories. We should also note that changes of greater magnitudes, alignment of actions, reassessing to determine effect of actions, and providing step-by-step examples for improvement can be extended beyond learning. The general principles could be applied to retention efforts, donor giving, or other important efforts. The more we discuss improvement, the better institutional decision makers we become.

Study Limitations and Future Directions


Thus far we have evaluated assessment reports from only one institution. Our findings might not reflect other institutions, especially those with different assessment practices and educational research initiatives. In addition, we have evaluated reports from only a single year's reporting cycle. Replicating our study across various reporting cycles, and across

institutions, would reveal potential longitudinal trends and could provide external validity evidence for our findings.

In addition, future research should include interviews with faculty members who crafted the assessment reports. Through these qualitative data, institutional effectiveness researchers could further investigate faculty perceptions of the magnitude of their described changes to curriculum and pedagogy. A rigorous qualitative follow-up study could also provide crucial insights from faculty members to clarify why certain types of information and explanations were absent from the reviewed assessment reports.

REFERENCES

- Banta, T. W., & Blaich, C. F. (2011, January). Closing the assessment loop. *Change: The Magazine of Higher Learning*, 22–27.
- 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.
- 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. & Bashkov, B. M. (2012, November). Do we practice what we preach? The accountability of an assessment office. *Assessment Update*, 24(6), 5–7, 14.



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.

Kinzie, J., Hutchings, P, & Jankowski, N.A. (2015). *Fostering greater use of assessment results: Principles for effective practice*. In G. D. Kuh, S. O. Ikenberry, N. A. Jankowski, T. R. Cain, P. T. Ewell, P. Hutchings, P., & J. Kinzie (Eds.), *Using evidence of student learning to improve higher education* (pp. 51–72). San Francisco, CA: Jossey-Bass.

Kuh, G. D., & Ikenberry, S. O. (2009). *More than you think, less than we need: Learning outcomes assessment in American higher education*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Kuh, G. D., Jankowski, N., Ikenberry, S. O., & Kinzie, J. (2014). *Knowing what students know and can do: Current state of student learning outcomes assessment in U.S. colleges and universities*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Jonson, J. L., Guetterman, T., & Thompson, R. J. (2014). An integrated model of influence: Use of assessment data in higher education. *Research and Practice in Assessment*, 9, 18–30.

Rodgers, M., Grays, M. P., Fulcher, K. H., &

Jurich, D. P. (2013) Improving academic program assessment: A mixed methods study. *Innovative Higher Education*, 38(5), 383–395.

Smith, K. L., Good, M. R., Sanchez, E. H., & Fulcher, K. H. (2015). Communication is key: Unpacking “use of assessment results to improve student learning.” *Research and Practice in Assessment*, 10, 15–29.

Suskie, L. (2010). Why are we assessing? *Inside Higher Education*. Retrieved from: <https://www.insidehighered.com/views/2010/10/26/suskie>

Taylor, P., Parker, K., Fry, R., Cohn, D., Wang, W., Velasco, G., & Dockterman, D. (2011). Is college worth it? College presidents, public assess value, quality and mission of higher education. Pew Research Center, Washington, D.C. Retrieved from <http://www.pewsocialtrends.org/2011/05/15/is-college-worth-it/>

Appendix A. Assessment Progress Template (APT) Evaluation Rubric as Described in Fulcher & Orem (2010)

Assessment Progress Template (APT) Evaluation Rubric

1 – Beginning	2 – Developing	3 – Good	4 – Exemplary
1. Student-centered learning objectives			
A. Clarity and Specificity			
No objectives stated.	Objectives present, but with imprecise verbs (e.g., know, understand), vague description of content/skill/or attitudinal domain, and non-specificity of whom should be assessed (e.g., "students")	Objectives generally contain precise verbs, rich description of the content/skill/or attitudinal domain, and specification of whom should be assessed (e.g., "graduating seniors in the Biology B.A. program")	All objectives stated with clarity and specificity including precise verbs, rich description of the content/skill/or attitudinal domain, and specification of whom should be assessed (e.g., "graduating seniors in the Biology B.A. program")
B. Orientation			
No objectives stated in student-centered terms.	Some objectives stated in student-centered terms.	Most objectives stated in student-centered terms.	All objectives stated in student-centered terms (i.e., what a student should know, think, or do).
2. Course/learning experiences that are mapped to objectives			
No activities/ courses listed.	Activities/courses listed but link to objectives is absent.	Most objectives have classes and/or activities linked to them.	All objectives have classes and/or activities linked to them.
3. Systematic method for evaluating progress on objectives			
A. Relationship between measures and objectives			
Seemingly no relationship between objectives and measures.	At a superficial level, it appears the content assessed by the measures matches the objectives, but no explanation is provided.	General detail about how objectives relate to measures is provided. For example, the faculty wrote items to match the objectives, or the instrument was selected "because its general description appeared to match our objectives."	Detail is provided regarding objective-to-measure match. Specific items on the test are linked to objectives. The match is affirmed by faculty subject experts (e.g., through a backwards translation).
B. Types of Measures			
No measures indicated	Most objectives assessed primarily via indirect (e.g., surveys) measures.	Most objectives assessed primarily via direct measures.	All objectives assessed using at least one direct measure (e.g., tests, essays).
C. Specification of desired results for objectives			
No a priori desired results for objectives	Statement of desired result (e.g., student growth, comparison to previous year's data, comparison to faculty standards, performance vs. a criterion), but no specificity (e.g., students will grow; students will perform better than last year)	Desired result specified. (e.g., our students will gain ½ standard deviation from junior to senior year; our students will score above a faculty-determined standard). "Gathering baseline data" is acceptable for this rating.	Desired result specified and justified (e.g., Last year the typical student scored 20 points on measure x. The current cohort underwent more extensive coursework in the area, so we hope that the average student scores 22 points or better.)
D. Data collection & Research design integrity			
No information is provided about data collection process or data not collected.	Limited information is provided about data collection such as who and how many took the assessment, but not enough to judge the veracity of the process (e.g., thirty-five seniors took the test).	Enough information is provided to understand the data collection process, such as a description of the sample, testing protocol, testing conditions, and student motivation. Nevertheless, several methodological flaws are evident such as unrepresentative sampling, inappropriate testing conditions, one rater for ratings, or mismatch with specification of desired results.	The data collection process is clearly explained and is appropriate to the specification of desired results (e.g., representative sampling, adequate motivation, two or more trained raters for performance assessment, pre-post design to measure gain, cutoff defended for performance vs. a criterion)
E. Additional validity evidence			
No additional psychometric properties provided.	Reliability estimates (e.g., internal consistency, test-retest, inter-rater) provided for most scores, although reliability tends to be poor (<.60). Or, author states how efforts have been made to improve reliability (e.g., raters were trained on rubric).	Reliability estimates provided for most scores, most scores are marginal or better (>.60).	Reliability estimates provided, most scores are marginal or better (>.60). Plus, other evidence given such as relationship of scores to other variables and how such relationship strengthens or weakens argument for validity of test scores.
4. Results of program assessment			

Appendix A continued on next page

Appendix A continued

A. Presentation of results			
No results presented	Results are present, but it is unclear how they relate to the objectives or the desired results for the objectives.	Results are present, and they directly relate to the objectives and the desired results for objectives but presentation is sloppy or difficult to follow. Statistical analysis may or may not be present.	Results are present, and they directly relate to objectives and the desired results for objectives, are clearly presented, and were derived by appropriate statistical analyses.
B. History of results			
No results presented	Only current year's results provided.	Past iteration(s) of results (e.g., last year's) provided for some assessments in addition to current year's.	Past iteration(s) of results (e.g., last year's) provided for majority of assessments in addition to current year's.
C. Interpretation of Results			
No interpretation attempted	Interpretation attempted, but the interpretation does not refer back to the objectives or desired results of objectives. Or, the interpretations are clearly not supported by the methodology and/or results.	Interpretations of results seem to be reasonable inferences given the objectives, desired results of objectives, and methodology.	Interpretations of results seem to be reasonable given the objectives, desired results of objectives, and methodology. Plus, multiple faculty interpreted results (not just one person). And, interpretation includes how classes/activities might have affected results.
5. Documents how results are shared with faculty/stakeholders			
No evidence of communication	Information provided to limited number of faculty or communication process unclear.	Information provided to all faculty, mode and details of communication clear.	Information provided to all faculty, mode and details of communication clear. In addition, information shared with others such as advisory committees, other stakeholders, or to conference attendees.
6. Documents the use of results for improvement			
A. Improvement of programs regarding student learning and development			
No mention of any improvements.	Examples of improvements documented but the link between them and the assessment findings is not clear.	Examples of improvements (or plans to improve) documented and directly related to findings of assessment. However, the improvements lack specificity.	Examples of improvements (or plans to improve) documented and directly related to findings of assessment. These improvements are very specific (e.g., approximate dates of implementation and where in curriculum they will occur.)
B. Improvement of assessment process			
No mention of how this iteration of assessment is improved from past administrations.	Some critical evaluation of past and current assessment, including acknowledgement of flaws, but no evidence of improving upon past assessment or making plans to improve assessment in future iterations.	Critical evaluation of past and current assessment, including acknowledgement of flaws; Plus evidence of some moderate revision, or general plans for improvement of assessment process.	Critical evaluation of past and current assessment, including acknowledgement of flaws; both present improvements and intended improvements are provided; for both, specific details are given. Either present improvements or intended improvements must encompass a major revision.

Thank You!

AIR expresses sincere appreciation for the members who serve as peer reviewers. Authors who submit materials to AIR Publications receive feedback from AIR members. The following individuals reviewed manuscripts submitted to this volume of *Professional Files*.

Matt Anson

Nora Galambos

Pa Thao

Dave Becher

Jennifer Kobrin

Benita Thompson

Anne Marie Brady

Ethan Kolek

Ying Vuthipadadon

René Cintrón

Teresa Isbell

Annie Weber

Amanda Clark

Jessica Lillegaard

Mike Williford

Linda Clark

Sun No

Liyong Cui

Gee Lockhart Sigman

About AIR *Professional File*

Professional File volumes are journal-length publications grounded in relevant literature that synthesize current issues, present new processes or models, or share practical applications related to institutional research. All submissions are peer-reviewed. For more information about AIR Publications, including *Professional File* and instructions for authors, visit www.airweb.org/publications.

Association for Institutional Research
Christine M. Keller, Executive Director and CEO
www.airweb.org

ISSN 2155-7535



AIR provides higher education professionals with learning opportunities that better equip them to meet the needs of data-informed decision making at their institutions. Whether online or face-to-face, in the office or at our annual conference, AIR's professional development opportunities provide outstanding return on investment for those working in IR and related fields, and the institutions they represent.

A Holistic Approach to IR

This course provides a foundation for participants to meet and navigate the ever-growing demands for data and information in the current higher education landscape.

IPEDS Keyholder Essentials: A Beginner's Guide

Created for data providers with less than 9 months of experience as keyholders, this course covers basic concepts and definitions.

IPEDS Keyholder Efficiencies: Reducing the Reporting Burden

Created for data providers with 10-24 months of experience, this course is designed to enhance your knowledge as an IPEDS Keyholder.

Data and Decisions Academy®

The Data and Decisions Academy is a series of online, self-paced courses designed for community college IR professionals.

AIR Forum

Learn, connect, and share at the largest gathering of IR professionals in the world, and take advantage of recorded sessions, presentation slides, handouts, and scholarly papers available after the conference.



LEARN MORE AT WWW.AIRWEB.ORG