OFFICE OF INSTITUTIONAL EFFECTIVENESS

ASSESSMENT OF ACADEMIC PROGRAMS

Primary Methodological Issues with Assessment Methods and Their Solutions

There are a number of persistent problems reflected in the undergraduate, graduate and certificate assessment procedures for which the Office of Institutional Effectiveness offer solutions in this document. While, two of the most common issues are the use of classroom grades as a measure of learning outcomes and an over-reliance on indirect measures as primary assessment tools, there are other important obstacles to well-developed assessment plans and methodologies. This brief paper provides guidance to departments in their efforts to avoid methodological problems with assessment approaches. The key code used in the assessment scorecard provides an organizational scheme for the solutions below.

Section Focusing on NSO

Distinguishing Learning Goals from Learning Outcomes

A common problem in assessment plans is the lack of distinction made between learning goals and learning outcomes. Goals are broader statements of intended learning and are usually general assertions regarding what will happen to students during a particular course or curriculum. Learning outcomes are more specific statements focusing on what students will be able to do or accomplish after a particular course, block of learning, or curriculum.

The two statements below are good examples of this distinction (these examples are from different USF assessment plans – names or references to departments have been removed):

Learning Outcome (NSO Example): Students will have adequate knowledge of Science

The example above provides an adequate learning goal, but it is too general a statement to be considered an adequate learning outcome. Contrast this example with the two below.

Learning Outcome (Good Practice Example 1): Upon completion of the core undergraduate courses, students will be able to demonstrate knowledge of and perform comprehensive data analysis to compose a comprehensive data analysis including interpretation of a data set.

Learning Outcome (Good Practice Example 2): Majors in this program will be able to conduct original research in this discipline using appropriate methods. They should demonstrate the ability to 1) formulate a meaningful hypothesis or research question; 2) clearly describe and justify the methods used; 3) present clear findings in written form.

The learning outcome statement below pertains to situations where students in a Certificate program are assessed as part of a degree program. Inclusion of this example is due to SAS recent requirement
of assessment of Certificate programs including those assessed within the context of a degree program.

Learning Outcome (Good Practice Example 1): Upon completion of the program, majors will be able to compare and contrast various forms of research methodologies including advanced quantitative methods. [This outcome applies to regular BS students as well as students pursuing a Certificate in Research Methods).

Section Focusing on GO, IDM, IR, and RV

Departments Using Course Embedded Assessment

- Capstone courses (IR)
  - Usually involve a capstone project
  - Assessment of capstone projects should have the following elements
    - **Multiple readers/scorers** (A minimum of two faculty members)
    - Clear and concise **rubrics**
    - A statement regarding a method for measuring **inter-rater reliability** for consistency of measurement.
  - Portfolios usually involve a combination of class work and assignments
  - Assessment of portfolios should have the following elements
    - **Multiple faculty scorers/reviewers**
    - Rubrics or common criteria for evaluating quality
    - A statement regarding a method for measuring **inter-rater/reviewer reliability** for consistency of measurement.

- Writing samples or Oral Presentations (IR)
  - Scoring should be conducted by **multiple or juried reviewers**
  - Clear and concise **rubrics** should be applied
  - A statement regarding a method for measuring **inter-rater reliability** for consistency of measurement.

- Common test item bank (RV)
  - Usually involves some bank of embedded items in tests which are either used consistently or rotated at random
  - Such items should have the following characteristics
    - Some effort to judge **reliability** (statistical methods to measure consistency are simplest)
    - Some effort to judge **validity** (relationship with other performance criteria)
    - OR selection of items from a test bank with already proven reliability and validity, e.g. text book banks that report psychometrics or “off-the-shelf” solutions (see section below for further information about “store-bought” measures
    - Item analysis so that appropriate item discrimination may be ascertained
    - IMPORTANT NOTE: embedded items should be considered a distinct part of the test which is focused on outcomes assessment and NOT a student grade.
Remember, assessment represents *formative evaluation of programs* not *summative evaluation* of students (implied by grades).

- **Indirect Measures (IDM)**
  - *Indirect measures* are usually surveys, focus groups, or interviews asking students to respond to prompts related to opinions or skill sets. For example, a student may be asked a level of agreement with a statement such as, “during my undergraduate coursework I feel I have become a better writer.”
  - There is a marked difference between a student’s report that s/he has become a better writer and a skill test to measure quality of writing.
  - Indirect measures usually require that faculty infer actual student abilities, knowledge, and values rather than observe direct evidence. *Indirect measures are generally viewed as a means of acquiring agreement between skills tests and student self-perceptions of skills and as such should not be used as the sole or predominant means of measuring learning outcomes.*

- **Grades as Outcomes Measures (GO)**
  - Accrediting bodies have increasingly eschewed the use of grades, e.g., “minimum grade of C,” as a valid measure of learning outcomes.
  - Therefore, *grades should never be the sole measure of a given learning outcome*. Again, grades represent a summative evaluation of a student NOT formative evaluation of a program.

**Solutions**

In assessment plans in USF Assessment Management System, the following will easily resolve most departmental methodological issues.

- State that *multiple reviewers* will be used. Include the number and method, e.g. two reviewers will score the portfolios; if the two reviewers are x scores apart, a third reviewer will act as mediator. The closest two scores will be used.
- Find a good *rubric* extant in the literature and use or modify it for your purposes. The Office of Institutional Effectiveness and Assessment is available to assist in the development of methodologically sound rubrics.
- Make some statement about evaluating *inter-rater reliability*, e.g. the percentage of agreement among reviewer will be calculated (in cases of categorical scales) OR a correlation between reviewer scores will be calculated (in cases of continuous scales. State that reviewers will be given some kind of training so that scoring schemes will become more consistent in the future. This training is referred to as *calibration* of the rubric.
  - SUGGESTED STATEMENT FOR IR (to be embedded in assessment plans in the appropriate place): “The department assessment coordinator will work with the Office of Institutional Effectiveness and Assessment to monitor inter-rater reliability.”
- **The use of direct measures, even if cyclic, should be used to validate student self-perceptions.**
Departments Using Off-the-Shelf Instruments

- Problems are reduced here because most of these instruments have been normed for target populations
- Still there can be problems in the way these are described in assessment plans
- Included in descriptions should be:
  - Correct name of instrument
  - Some *brief* statement about *reliability and validity* of the instrument
  - Sampling procedures for student testing including number of students, estimated percentage of entire population represented by the sample, and time of testing, e.g. rising juniors, graduating seniors, etc.
  - Local standardization should be used initially to check for applicability to the student population at the institution using the instrument.

**Section Focusing on IDF, ISE, and MOD**

These problems are sometimes less obvious but no less toxic to a good assessment plan. They are so similar that the differences between them are merely a matter of emphases in an assessment report. Even reports with clearly stated learning outcomes and measures may not connect all the dots in what accrediting bodies call *closing the loop*. This section elaborates further on the definitions provided in the “Key Code” for assessment scorecards. The examples used in this section are based on real cases reflected in the *USF Assessment Management System*.

**IDF:** An incomplete data feedback loop occurs when a measure and/or finding in the *measures and findings* section is not stated with sufficient specificity for the reviewer to infer the means by which program improvement can occur from the measure and finding.

For example, several learning outcomes may be “batch measured” by a single portfolio. However, if the portfolio description bears no clear relation to the program goals and outcomes, its relevance as a measure used to “close the loop” is lost.

**IDF Solution:** Make sure that in the description of the measure as well as in the presentation of findings, there is a clear connection made to the mission and goals of the program. Statements about changes made to the pedagogical cycle and/or pedagogical style can be particularly robust here. For example, if the findings from the analysis of portfolios have lead to a change from, say, the lecture format to a case-study format, the case for “closing the loop” has become very powerful.

**ISE:** An incomplete source of evidence error occurs when a measure appears to be inappropriate or so vaguely stated that its relationship to a learning outcome is suspect.

Some of these cases are more obvious than others. For example, using an indirect measure, e.g., student survey, as the sole source to evaluate learning of course content is an obviously insufficient
source of evidence. However, it is less clear whether or not embedded test items are sufficient unless there is a complete discussion of these items to include the kind of items (are they multiple choice or some other type), number of items per outcome, and efforts to evaluate the reliability and validity of the items.

An oft repeated example of ISE in USF Assessment Management System is using “completion of course X” as a measure of a learning outcome. Others include institutional initiatives such as faculty or TA training, attendance at conferences or other events, and description of course content as the measure for an outcome.

**ISE Solution:** Thoroughly describe the nature of measures and their relationship to outcomes. Include efforts to validate measures (this error often goes hand-in-hand with an RV error) as well as changes to the measures based on their utility or psychometric properties. Avoid inappropriate measures for learning outcomes such as a course offering.

**MOD:** Whereas, an IDF error is focused on the feedback between an outcome or measure and program improvement, a measure(s) disconnected from outcomes (MOD) error occurs when the measures used to evaluate student progress on a given outcome are insufficiently precise to make an obvious connection between the outcome and a measure.

For example, let’s say a given plan in USF Assessment Management System contains two learning outcomes, one focusing on critical thinking and one targeting written communication skills. However, in the measures and findings section, a single measure, in this case, a rubric is put forth as the measure for both critical thinking and written communication. In some cases, the plan fails to provide sufficient description of the rubric, and the connection between the rubric and both learning outcomes is not made.

**MOD Solution:** Make sure measures are described with sufficient specificity and that each learning outcome referenced has a scale inherent in the measure. Provide copies of rubrics by uploading all relevant rubrics into the document repository and associating them with relevant learning outcomes.