

Assessment for the New Curriculum: A Guide for Professional Accounting Programs

Section 9.2 Measuring Knowledge Outcomes

It is often tempting to initiate an assessment program by administering a standardized, machine-scored test of knowledge to seniors. This strategy has several limitations:

FIGURE 9.2
COMPARING STANDARDIZED AND LOCALLY DEVELOPED
ACHIEVEMENT EXAMINATIONS

STANDARDIZED EXAMS	
Advantages	Disadvantages
<ul style="list-style-type: none"> • lower opportunity cost to obtain and administer test • high level of technical quality (valid, reliable) • reflect expert views of what students should know • national norms allow comparison of individuals and/or groups or to an absolute standard (criterion) • quick and convenient if available for departmental use 	<ul style="list-style-type: none"> • limited options to choose from • test content may not be consistent with program objectives • test content may lack breadth of content and scope of skills measured • comparative testing communicates competitive ethic • forced-choice testing communicates low priority on creativity • availability may be limited to public testing sessions • low student motivation to take test if no personal consequences or intrinsic value
LOCALLY DEVELOPED EXAMS	
Advantages	Disadvantages
<ul style="list-style-type: none"> • can be designed to match program objectives • scope and length of test can be adjusted to departmental needs • faculty involvement fosters dialogue about educational ends and means • variety of test formats can be used • testing can be incorporated into classroom instruction, maximizing student participation and motivation 	<ul style="list-style-type: none"> • high opportunity cost to develop and validate test • faculty may not agree on test content • no point of comparison beyond institution • test security requires annual updating or large test item pool

- Standardized examinations may emphasize the lower-level objectives of Bloom's taxonomy ([Chapter 6](#)) rather than the current expanded definition of knowledge as the ability to apply and adapt concepts and principles.
- Testing with standardized instruments yields useful diagnostic information only if the test specifications correspond closely to program objectives, and if group subscores keyed to program objectives can be obtained.
- Results of a senior exit test cannot be used to benefit the students who take it, defeating a central purpose of assessment. Experience indicates (Jacobi, Astin, and Ayala, 1987; Johnson, 1993) that many students simply decline to participate in the assessment, regardless of incentives offered.

This section briefly summarizes procedures for developing or selecting knowledge outcome measures and analyzes options for obtaining information on knowledge outcomes. It also suggests ways to improve the usefulness of regular course examinations for program-level assessment.

9.2.1. Procedures for Developing Knowledge Outcome Measures

Expanded knowledge outcomes, as well as professionally-relevant skills, can be assessed using traditional "paper-and-pencil" measures and "performance" assessments. This section briefly summarizes procedures for the more traditional and familiar measures, while Section 9.3 focuses on performance assessment with emphasis on measuring intellectual, communication, and interpersonal skills.

The selection or development of knowledge outcome measures begins with a set of *objectives and performance criteria* as described in [Chapter 7](#). The objectives should address the full range of *uses* of knowledge desired by the faculty, consistent with the expanded definition of knowledge needed in the new professional environment

Using objectives and performance criteria greatly simplifies the task of designing measures. For knowledge outcomes, objectives will be translated into "objective" or "forced-choice" test items and "open-ended" or "free-response" items (for details of item construction and test design, see McBeath, 1992). While in principle objective tests can measure high-level cognitive outcomes, in practice it may be difficult to construct good forced-choice items to test these outcomes (Banta and Schneider, 1988). Fortunately, as noted earlier, some measures already used by faculty in their courses may be adaptable for purposes of program-level assessment. Sample materials for the full spectrum of knowledge and skill objectives can be found in curriculum resources such as the BYU report (1992) and USC's *Year 2000 Curriculum Project* (Diamond and Pincus, 1994; Pincus, 1993), as well as in periodicals such as the *Journal of Accounting Education* and *Issues in Accounting Education*. Section 9.2.2 describes and evaluates various measurement options for knowledge outcomes.

When open-ended items are used, student responses should be rated on *performance criteria* developed by the faculty for the targeted objective(s). The development of performance criteria is described in Section 9.3. Faculty who use performance ratings should participate in rater training (also described in Section 9.3), whether they are applying the criteria for a course- or program-level evaluation.

Field testing or pilot testing is useful to prevent problems of unclear or inappropriate items or excessive test length. The instrument can be tested on a small group of students who have previously completed the relevant coursework and who would therefore not be part of the assessment sample. The validation sample should be similar to the group for which the measure is designed. No instrument should be used for the first time in a large-scale, high-stakes assessment.

Analyzing *reliability and validity* helps to interpret results of the pilot test and to determine whether the findings can be safely used to make generalizations about students' strengths, weaknesses, and overall performance on the objectives measured.

Modifications of the instrument will be suggested by analysis of pilot test results, reliability/validity data, discussion of results by the faculty, and review of students' feedback about the measure (obtained through discussion with students or using a brief questionnaire).

9.2.2 Options for Measuring Knowledge Outcomes

Options for measuring knowledge range from familiar course-embedded measures to formal achievement examinations and self-report data. By using a cluster of measures, faculty can balance the advantages and disadvantages of a single measure to obtain an overall portrait of students' strengths and weaknesses.

- *Course examinations and project grades* provide useful data to the extent these measures address targeted program objectives and produce subscores related to those objectives. This requirement can be met by compiling results for all students on each objective to create a

profile of students' accomplishments.

The value of examination scores for assessing aggregate performance of students increases when a common examination or item pool is used by all who teach the course, and when the test is demonstrated to be acceptably reliable and valid.

- *Course grades*: Grades in individual courses have limited usefulness as outcome measures because the basis for grading is not usually standardized. Grades include elements besides performance on knowledge tests, ranging from attendance and class participation to scores on group projects. Course grades can be made more informative as measures of learning outcomes by ensuring that they are based solely or primarily on students' performance relative to a specified subset of program objectives.
- *Grade-point averages (GPA)*: Overall GPA is considered a moderately useful outcome measure by accounting program administrators (Chamberlain and others, 1991) and is widely used by employers of accounting graduates as well, as the following comment suggests:

Despite all the drawbacks of individual course grades, overall GPA does relate to job performance (and probably graduate school performance). It is a multi-year, multi-method, multi-rater measure and when used across institutions in the range we hire from, it works. (Jean Wyer, personal communication, March 1994)

Although GPA "works" as a guide for employers, it is not useful as a diagnostic indicator for individual students nor does it provide useful information about program quality. GPA is a useful predictor for employers because of a partial overlap between the knowledge and skills assessed by some faculty and those needed for successful employment. GPA may also be a proxy indicator of non-academic skills that are needed for professional practice, for example, aggressiveness, the ability to navigate bureaucracies, and ability to organize one's time effectively (Peter Ewell, personal communication, March, 1994). The predictive validity of the GPA can be expected to be even greater when it reflects systematic assessment of students' performance on knowledge, skills, and values identified as central to professional success.

- *Achievement examinations* are designed to provide a cumulative assessment of knowledge retained by students. A risk is that the instrument may over-emphasize lower-level cognitive objectives for the sake of administrative convenience. As noted earlier, if the examination is given at the end of the senior year, the results are of little value to the students who take it. An achievement test designed to be administered late in the junior year can provide diagnostic information for both the program and students, again using the strategy of building the test around program objectives and providing subscores for each objective.

When using a standardized exam, test specifications should be examined closely to determine the degree of overlap with program objectives. Existing tests include:

CPA Exam: Although pass rates on the CPA exam are frequently used as an indicator of program quality, it is a licensure exam for public accounting, not a test of general achievement in accounting education. It is time-consuming and expensive, and because it is not required in order for graduates to practice accounting, self-selection bias precludes making judgments of program quality based on results. Moreover, it cannot be taken until students have graduated, nor does it provide subscores, so it cannot be used to provide timely diagnostic of students' strengths and weaknesses or to identify areas in the curriculum that may need attention (Herring and Izard, 1992). Thus, like GPA, the results may be useful for employers but have minimal value for program improvement.

Achievement Test for Accounting Graduates (ATAG): This examination (formerly the AICPA Level II Achievement Test) yields subscores in five areas (auditing, financial, cost and managerial, accounting information systems, and taxation) that correspond to basic divisions in the accounting curriculum at most institutions (Herring and Izard, 1992). However, Ingram and Peterson (1987) found that AICPA exam scores did not improve the predictive ability of regression models based on ACT scores and grades in lower-division accounting courses.

Other standardized exams: Other available instruments include the ACT Proficiency

Examination Program in Accounting, designed for awarding credit by examination. The exam tests accounting proficiency at three levels, using objective and essay tests. ETS offers a course equivalency examination on Principles of Financial Accounting, designed for nontraditional students (Smith, Draper, and Bradley, 1994).

For more general assessment of business knowledge, the AACSB offers the Core Curriculum Assessment Program (CCAP). The CCAP is a data base which includes questions on accounting as well as other areas in the business curriculum. Schools may purchase the database on diskette and use it to design customized examinations. (Baker and other, 1993).

Instructor ratings: Checklists or narratives can be used to obtain judgments of students' knowledge and skills or of their performance in specific areas. While such judgments are subject to a number of sources of error (for example, "halo effects"), they offer a convenient way to obtain an estimate of students' strengths and weaknesses with respect to a particular category of outcomes.

Ratings are more reliable when they are based on specific performances such as oral or written examinations, presentations, projects, and simulations, and when the instructor or other rater has been trained to use agreed-upon performance criteria, as discussed in Section 9.3.

Student self-reports: Self-reports can provide a convenient profile of students' self-assessed strengths and weaknesses of both knowledge and skills. They can be used to gauge learning from specific instructional methods and materials (such as the multimedia resource, *Dermaceutics Inc.:* Risk Assessment and Planning; see example, [Appendix 1](#)). They can also be used to assess broad curricular outcomes. For example, several schools have devised questionnaires which ask seniors and alumni to judge their current level of knowledge and skill, based on AECC, FSA, or departmental objectives. [Appendix 2](#) presents an example of a student self-report form from Arizona State University.

Advantages and Limitations: Self- and instructor-ratings have two important advantages: they can be obtained quickly and inexpensively, and they can be used to assess all major categories of learning outcomes: knowledge, skills, and values and attitudes.

Self-ratings are moderately correlated with other measures such as standardized examination scores but are not sufficiently valid to be used without other corroborating measures (Ewell, 1993). The problem is illustrated in an assessment of the *Dermaceutics* package mentioned above. Students in an experimental group studied audit concepts using the *Dermaceutics* package, while students in the control group participated in case discussions of extra problems and questions from the text. Experimental and control students did not differ significantly on *self-reported* understanding of the concepts. However, the actual *performance* of students in the experimental group was superior to that of the control group on multiple-choice and essay questions included on a mid-term examination (Mohrweis, 1993).

Because their validity is limited, self-report data should not be used to make placement or proficiency decisions, or other formal judgments about individual students. They are probably most useful for identifying potential problem areas. Checklists and other self-report measures may also facilitate students' discussions with their advisors about ways to improve their performance.

9.2.3 Improving the Quality of Knowledge Measures

Knowledge outcome measures can often be improved by following a few simple guidelines. Here are nine; others can be found in resources such as McBeath (1990), Erwin (1991), Ball State (n.d.) and references cited therein.

1. Prepare a test map indicating the objectives to be addressed and the weight given to each. Use to determine the number of items for each objective and points assigned to each.
2. Include full range of knowledge outcomes in test specifications.
3. When using standardized tests, identify areas of overlap with program objectives as well as gaps in test coverage; compensate for the gaps and avoid drawing conclusions based on results for topics covered on the test but not included in the curriculum.

4. Observe basic rules for test item writing (for example, avoid double negatives, "all" or "none of the above," item stems that are too short or too vague).
5. Run a test analysis program to uncover weak items on objective tests.
6. Use strategies to minimize error when scoring essay tests, for example:
 - Develop a scoring key; see, for example, Scofield and Combes (1993), reproduced in [Appendix 3](#).
 - Rate all responses to a single item in one pass, then shuffle the papers and rate all responses to the second item, and so on (to reduce contamination of scores and effects of order of presentation).
 - Rate essays "blind," that is, ask students to put their names on the back of the paper or use a code.
7. Ask a colleague to review a draft of the test; see if your answers agree, and check for confusing wording.
8. Pilot test the instrument on a small sample of students; modify if necessary.
9. Use multiple measures to obtain a comprehensive picture of students' understanding of accounting concepts and principles and their use.