

Possible Assessment Methods for use with the Student Learning Outcomes

1. Alumni Surveys
2. Archival Data
3. Capstone Project
4. Culminating Assignments
5. Content Analysis
6. Course Embedded Assessment
7. Curriculum Analysis
8. Delphi Technique
9. EPortfolio
10. Employer Surveys
11. Focus Groups
12. Institutional Data
13. Matrices
14. Observations
15. Oral Exams
16. Performance Assessment
17. Portfolio Evaluations
18. Pre-test/Post-test Evaluations
19. Reflective Essays
20. Rubrics
21. Standardized and Local Test Instruments
22. Student Self-efficacy
23. Student Surveys and Exit Interviews
24. Syllabus Analysis
25. Transcript Analysis
26. Value Added Assessment (Pre- and Post-testing)

Alumni Surveys

Description: The perspective that students have on their education may change significantly after time away from school. Some learning outcomes lend themselves more naturally to questions posed some time after graduation. Surveying alumni can provide a wide variety of information about program satisfaction, how well students are prepared for their careers, what types of jobs or graduate degrees majors have gone on to obtain, starting salaries for graduates, and the skills that are needed to succeed in the job market or in graduate study. These surveys provide the opportunity to collect data on which areas of the program should be changed, altered, improved or expanded.

Strengths and Weaknesses: Alumni surveying is usually a relatively inexpensive way to collect program data from individuals who have a vested interest in helping you improve your program as well as offering the opportunity for improving and continuing department relationships with program graduates. However, without an easily accessible and up-to-date directory of alumni, they can be difficult to locate. The response rate may be very low and the results may represent bias populations. It also takes time to develop an effective survey and ensure an acceptable response rate.

Adapted from Palombo et al. Ball State University, Assessment Workbook (2000).

Additional Resources:

Converse, J. M. & Pressler, S. (1986). *Survey questions: Handcrafting the standardized questionnaire*. SAGE Publications.

Dillman, D. (1978). *Mail and telephone surveys: The total design method*. New York: Wiley-Interscience Publication.

Dyke, J. V. & Williams, G. W. (1996). Involving graduates and employers in assessment of a technology program. In Banta, T. W., Lund, J. P., Black, K. E., & Oblander, F. W. (Eds.). *Assessment in practice*, pp. 99-101. San Francisco: Jossey-Bass Publishers.

Ewell, P. (1983). *Student outcomes questionnaires: An implementation handbook*. New York, NY: National Center for Higher Education Management Systems and the College Board.

Labaw, P. J. (1980). *Advanced questionnaire design*. Cambridge, MA: Abt Books.

McKenna, B. *Surveying your alumni: Guideline and 22 sample questionnaires*. Washington, DC: Council for advancement and support of education.

Archival Data

Description: Biographical, academic, or other file data available from archive systems. Primarily aggregated student information but in some cases can be individual student data. Can use comparable data from other institutions for benchmarking.

Can be quick, easy and cost-effective method if data is available and accessible. Integral to valuable longitudinal comparisons. Should be a standard component of all assessment programs.

Strengths and Weaknesses:

Advantages:

- Tend to be accessible, thus requiring less additional effort.
- Build upon efforts that have already occurred.
- Can be cost efficient if required data is readily retrievable in desired format.
- Constitute unobtrusive measurement, not requiring additional time or effort from students or other groups.
- Very useful for longitudinal studies
- Ideal way to establish a baseline for before and after comparisons

Disadvantages:

- Especially in large institutions, may require considerable effort and coordination to determine exactly what data are available campus-wide and to then get that information in desired format.
- To be most helpful, datasets need to be combined. This requires an ability to download and combine specific information for multiple sources. It may require designing a separate database management system for this downloaded information.
- Typically the archived data are not exactly what is required, so that the evaluator must make compromises. In some cases, it may be a stretch to use such data as surrogates for the desired measures.
- If individual records are included, protection of rights and confidentiality must be assured; should obtain Institutional Review Board approval if in doubt.
- Availability may discourage the development of other, more responsive measures or data sources.
- May encourage attempts to “find ways to use data” rather than measurement related to specific goals and objectives.

Ways to Reduce Disadvantages:

- Early-on in the development of an assessment program, conduct a comprehensive review of existing assessment and evaluation efforts and data typically being collected throughout the institution and its units (i.e., “campus data map”) - Institutional Research Office is a good place to start if you don’t maintain the archived data yourself.
- Be familiar with the Family Educational Rights and Privacy Act (Buckley Amendment) and avoid personally identifiable data collection without permission. Assure security/protection of records.
- Only use archival records that are relevant to specific goals and objectives of learning and development.

Additional Resources:

Astin, Alexander W. "Involvement in Learning Revisted: Lessons We Have Learned." *Journal of College Student Development*; v37 n2 p. 123-34, March 1996.

Astin, Alexander W.; et al., *Degree Attainment Rates at American Colleges and Universities: Effects of Race, Gender, and Institutional Type*. Higher Education Research Inst., Inc., Los Angeles, CA, 1996.

Culminating Assignments

Description: Culminating assignments offer students the opportunity to put together the knowledge and skills they have acquired in the major, provide a final common experience for majors, and offer faculty a way to assess student achievement across a number of discipline-specific areas. Culminating assignments are generally designed for seniors in a major or field to complete in the last semester before graduation. Their purpose is to integrate knowledge, concepts and skills that students are expected to have acquired in the program during the course of their study. This is obviously a curricular structure as well as an assessment technique and may consist of a single culminating course (a "capstone" course) or a small group of courses designed to measure competencies of students who are completing the program. A senior assignment is a final culminating project for graduating seniors such as a performance portfolio or a thesis that has the same integrative purpose as the capstone course.

Strengths and Weaknesses: Many colleges and universities are using capstone courses to collect data on student learning in a specific major or in general education or core requirement programs. Putting together an effective and comprehensive capstone course can be a challenge, however, particularly for those programs that mesh hands-on technical skills with less easily measurable learning outcomes. Also, there is a great deal of start-up time to developing appropriate and systematic methods for assessing these or other culminating experiences. See Content Analysis and Primary Trait Analysis below for further information.

Adapted from the University of Wisconsin, Madison, Outcomes Assessment Manual (2000)

Additional Resources:

Julian, F. D. (1996). The capstone course as an outcomes test for majors. Banta, T. W., Lund, J. P., Black, K. E., & Oblander, F. W. (Eds.). In *Assessment in practice*, pp. 79-81. San Francisco: Jossey-Bass Publishers.

Upcraft, M. L., Gardner, J. N., & Associates. (1989). *The freshman year experience: Helping students survive and succeed in college*. San Francisco: Jossey-Bass Publishers.

Content Analysis

Description: Content analysis is a technique that looks at a group of students, such as majors in a program or department, and assesses samples of written work that are produced by this group. This assessment method uses outcomes identified as important prior to the analysis or as the analysis proceeds. For example, you might want

to determine how well majors in your department write. To use content analysis to assess their writing skills, you will need a representative sample of the writing. Analysis may look at what students actually write or at the underlying meaning of their writing. Results are generally presented in written form giving averages and examples of specific categories of outcomes (e.g., spelling errors). Primary trait analysis, which identifies important characteristics of specific assignments and assigns levels of competency to each trait, can be particularly effective in identifying student learning.

Strengths and Weaknesses: Content analysis allows you to assess learning outcomes over a period of time and can be based on products that were not created for program assessment purposes. Because writing samples can be re-examined, content analysis also makes it easier to repeat portions of the study and provides an unobtrusive way to assess student learning. However, accuracy of the assessment is limited to the skill of the person(s) doing the analysis. Data is also limited by the set of written work and may not be relevant to technical skills valued by a particular field or major that involve hands-on performance. Pre-testing coding schemes, using more than one analyst per document, and concrete materials and coding schemes can improve the reliability of this technique.

Adapted from the California State University Bakersfield, PACT Outcomes Assessment Handbook (1999).

Additional Resource:

Babbie, E. (1995). *The Practice of Social Research (7th ed.)*. Belmont, CA: Wadsworth.

Walvoord, B. E. & Anderson, V. J. (1998). *Effective grading: A tool for learning and assessment*. San Francisco: Jossey-Bass.

Course-embedded Assessment

Description: Course-embedded assessment refers to methods of assessing student learning within the classroom environment, using course goals, objectives and content to gauge the extent of the learning that is taking place. This technique generates information about what and how students are learning within the program and classroom environment, using existing information that instructors routinely collect (test performance, short answer performance, quizzes, essays, etc.) or through assessment instruments introduced into a course specifically for the purpose of measuring student learning.

Strengths and Weaknesses: This method of assessment is often effective and easy to use because it builds on the curricular structure of the course and often does not require additional time for data collection since the data comes from existing assignments and course requirements. Course-embedded assessment does, however, take some preparation and analysis time and, while well documented for improving individual courses, there is less documentation on its value for program assessment.

Adapted from the University of Wisconsin, Madison, Outcomes Assessment Manual -2000, and the California State University, Bakersfield, PACT Outcomes Assessment Handbook-1999.

Additional Resources:

Angelo, T. A. & Cross, K. P. (1993). *Classroom assessment techniques: A Handbook for college teachers (2nd. Ed.)*. San Francisco: Jossey-Bass.

Classroom Assessment Techniques. (1999). Center for Excellence in Learning & Teaching. www.personal.psu.edu/celt/CATs.html

Palomba, C. A., & Banta, T. W. (1999). *Assessment essentials*. San Francisco: Jossey-Bass.

Walvoord, B. E. & Anderson, V. J. (1998). *Effective grading: A tool for learning and assessment*. San Francisco: Jossey-Bass.

Curriculum Analysis

Description: Accrediting bodies have historically required institutions to document the information that students are receiving and the content that the program delivers. With the move toward learning-outcomes assessment, programs are required to show that students actually exhibit the skills and qualities that the program wishes to develop. Curriculum analysis may be relevant. Curriculum analysis involves a systematic review of course syllabi, textbooks, exams, and other materials to help you clarify learning objectives, explore differences and similarities between course sections, and/or assess the effectiveness of instructional materials. It offers a way to document which courses will cover which objectives and helps in sequencing courses within a program. Also see Matrices.

Strengths and Weaknesses: Using curriculum analysis as an assessment tool can be a valuable way of tracking what is being taught where. It can provide assurance that specific learning goals and objectives are being covered in the program and can pinpoint areas where additional coverage is needed. This method, however, can be time-consuming, particularly in large departments with many courses and different instructors, and there may be little consistency between how learning objectives are addressed in one course and how they are taught in another.

Adapted from the Ball State University, Assessment Workbook, 1999 and The University of Wisconsin, Madison, Outcomes Assessment Manual 2000), the Indiana University Assessment Handbook, 2004.

Additional Resources:

Bers, T., Davis, D., & Taylor, W. (1996, Nov.-Dec.). Syllabus analysis: What are you teaching and telling your students? *Assessment Update* (8), 6, pp. 1-2, 14-15.

Diamond, R. M. (1998). *Designing and assessing courses and curricula*. San Francisco: Jossey- Bass.

Ewell, P. T. (1997). Identifying indicators of curricular quality. In *Handbook of the undergraduate curriculum*, J. G. Gaff & J. L. Ratcliff (Eds.). San Francisco: Jossey Bass, pp. 608-627.

Delphi Technique

Description: The Delphi technique is used to achieve consensus among differing points of view. In its original form, a team of experts, who never actually meet, are asked to comment on a particular issue or problem. Each member's response is reviewed and a consensus determined. Any member whose response falls outside of the consensus is asked to either defend or rethink the response. The anonymity provided by this technique offers more junior members of the team an equal chance to get their ideas out, as well as permits a challenge to the ideas of senior members that might never take place in an open forum. More recently, the Delphi technique has been modified so that teams of individuals are brought together to discuss an issue or problem face-to-face and reaching a consensus at the meeting. For instance, a team of faculty members might meet to review possible goals and objectives for their department in an effort to develop a set of goals and objectives on which they can agree.

Strengths and Weaknesses: The Delphi technique can be useful in bringing together diverse opinions in a discussion forum. This technique fails, however, when the facilitator lacks objectivity or when the participants feel unsafe or insecure in voicing their real opinions. For instance, a faculty member discussing intended goals and objectives might not be comfortable in disagreeing with the department head. For this technique to succeed, care must be taken to appoint an impartial facilitator and to convince participants that differing opinions are welcome. Returning to the original design of this technique, with an anonymous team who never meet, might ensure more honest and open input.

Additional Resources:

Armstrong, M. A. (1989). The Delphi technique. Princeton Economic Institute.

<http://www.pei-intl.com/Research/MARKETS/DELPHI.HTM>.

Cline, Alan. (2000). Prioritization Process using Delphi Technique.

www.carolla.com/wp-delph.htm.

Stuter, L. M. (1996). The Delphi technique: What is it?

<http://www.icehouse.net/lmstuter/page0019.htm>.

Stuter, L. M. (November 1998). Using the Delphi technique to achieve consensus. *Education Reporter* (54).

E-Portfolios

Description: "An E-Portfolio is a collection of electronic evidence assembled and managed by a student, usually on the web. Such electronic evidence may include inputted text, electronic files, images, multimedia,

blog entries and hyperlinks. E-Portfolios are both demonstrations of the user's abilities and platforms for self expressions" (Wikipedia)

E-Portfolios are created by the three principal activities of collection, selection and reflection, and are succinctly defined as collections of works selected from a larger archive of work on which the student has reflected.

Strengths and Weaknesses:

Advantages to the use of E-Portfolio assessment include:

- Drawn from the student's ongoing work (implications for validity, complexity, student motivation),
- Enhancement of learning (engagement, integration)
- Can be reliably scored with rubrics
- Can be structured to produce comparative information

Documenting learning in this way places the focus on actual achievements that are viewed directly, rather than on proxies of achievement like cumulative GPAs or test scores. The focus is on what students can DO with their knowledge and skills and not simply on whether knowledge has been acquired.

Disadvantages include:

- Different disciplinary models and uses for E-Portfolios
- Lack of longitudinal data on students or institutional use of E-Portfolios
- Software issues: designed for different purposes/uses of E-Portfolios; software infancy
- Extensive planning is needed

Possible uses of E-Portfolios for assessment include:

- Personal development planning
- Development of civic-minded students
- Leadership development
- Professional ethics
- Integrated learning

Additional Resources: <http://electronicportfolios.org/portfolios/bookmarks.html#hied>

Employer Surveys

Description: It is possible that some of the students' knowledge and skills are evident to the employers who rely on these characteristics. Thus, some accrediting bodies either require or encourage programs to perform an assessment through the major employers of their students. These may range from information as basic as hiring data, to site supervisor evaluations, to detailed surveys of the characteristics that the employers perceive in

program graduates. Employer surveys help determine if graduates have the necessary job skills and if there are other skills that employers particularly value that graduates are not acquiring in the program. This type of assessment method can provide information about the curriculum, programs and student outcomes that other methods cannot: on-the-job, field-specific information about the application and value of the skills that the program offers.

Strengths and Weaknesses: Employer surveys provide external data that cannot be replicated on campus and can help faculty and students identify the relevance of educational programs, although, as is true in any survey, ambiguous, poorly-worded questions will generate problematic data. Additionally, though data collected this way may provide valuable information on current opinion, responses may not provide enough detail to make decisions about specific changes in the curriculum or program. Also, it is sometimes difficult to determine who should be surveyed, and obtaining an acceptable response rate can be cost and time intensive.

Adapted from the Ball State University, Assessment Workbook (1999), the California State University, Bakersfield, PACT Outcomes Assessment Handbook (1999), and the University of Wisconsin, Madison, Outcomes Assessment Manual I (2000), the Indiana University Southeast Assessment Handbook, 2004)

Additional Resources:

Converse, J. M. & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire*. Newbury Park: SAGE Publications.

Dyke, J. V., & Williams, G. W. (1996). Involving graduates and employers in assessment of a technology program. In Banta. T. W., Lund, J. P., Black, K. E., & Oblander, F. W. (eds.) *Assessment in Practice*. San Francisco: Jossey-Bass.

Lead Center, University of Wisconsin, Madison. (1998). Program assessment toolkit: A guide to conducting interviews and surveys.

Focus Groups

Description: Focus groups are structured discussions among homogeneous groups of 6-10 individuals who respond to specific open-ended questions designed to collect data about the beliefs, attitudes and experiences of those in the group. This is a form of group interview where a facilitator raises the topics for discussion and collects data on the results. Emphasis is on insights and ideas.

Strengths and Weaknesses: Focus groups can provide a wide variety of data about participants' experiences, attitudes, views and suggestions, and results can be easily understood and used. These groups allow a small number of individuals to discuss a specific topic in detail, in a non-threatening environment. Data collected in this way, however, is not useful for quantitative results, and qualitative data can be time-consuming and difficult

to analyze because of the large amount of non-standardized information. Ultimately, the success of this method depends on a skilled, unbiased moderator and appropriate groups of participants.

Adapted from Palombo et al. Ball State University, Assessment Workbook (2000); and the California State University, Bakersfield, PACT Outcomes Assessment Handbook (1999).

Additional Resources:

Lead Center, University of Wisconsin, Madison. (1998). Program assessment tool kit: A guide to conducting interviews and surveys. Morgan, D. L. (1988). *Focus groups as qualitative research*. Newbury Park: SAGE Publications.

Morgan, D. L., & Krueger, R. A. (1997). *The focus group kit (Vols. 1-6)*. Thousand Oaks, CA: SAGE Publications.

Institutional Data

Description: A variety of departmental and student data are routinely collected at the university level. These data can enhance and elaborate on data you collect in the department. Institutional data can tell you whether the program is growing, what the grade point average is for majors in the program, and what the retention rate is for your students.

Strengths and Weaknesses: Institutional data are generally easily accessible and readily available. On the Purdue WL campus, you can access this data through the Office of Institutional Research (OIR), the ODS, the historic mainframe SQL server, DSS, and a variety of other locations. Student and departmental data are collected on a systematic and cyclical schedule that can offer you both current and longitudinal information. On the other hand, these data sets are generally large and may be difficult to sort through, particularly for those individuals who are not used to working through large databases.

The data may be less useful to specific departments or programs because the information collected is very often general (age, gender, race, etc.) and may not directly relate to program goals and objectives.

Adapted from the Ball State University, Assessment Workbook-1999.

Matrices

Description: At its most basic, a matrix is a grid of rows and columns used to organize information. For assessment purposes, a matrix can be used to summarize the relationship between program objectives and course syllabus objectives, course assignments, or courses in a program or department. Matrices can be used for curriculum review, to select assessment criteria or for test planning. A matrix can also be used to compare program outcomes to employer expectations.

Strengths and Weaknesses: Using a matrix can give you a good overview of how course components and curriculum link to program objectives, can help you tailor assignments to program objectives, and can lead to useful discussions that in turn lead to meaningful changes in courses or curricula. However, because a matrix can offer a clear picture of how program components are interconnected and can reveal where they are not, acknowledging and responding to discrepancies may involve extensive discussion, flexibility and willingness to change.

Adapted from the Ball State University, Assessment Workbook, revised April (2000), and the California State University, Bakersfield, PACT Outcomes Assessment Handbook (1999).

Additional Resource:

Diamond, R.M. (1998). *Designing and assessing courses and curricula*. San Francisco: Jossey-Bass.

Palomba, C. A., & Banta, T. W. (1999). *Assessment essentials*. San Francisco: Jossey-Bass.

Observations

Description: Observation as a method of assessment is an unobtrusive tool that can yield significant information about how and why students learn. You may choose to observe any relevant interactive event, such as classes, club meetings, or social gatherings. This tool is generally used when you are interested in how students study, are concerned about the effectiveness of study sessions or other supplementary activities, or when you are focusing on the relationship between out-of-class behavior and in-class performance. Data collected through observation can be correlated with test scores and/or course grades to help provide further insight into student learning.

Strengths and Weaknesses: Data collected through observation can yield important insight into student behavior that may be difficult to gauge through other assessment methods. This method is typically designed to describe findings within a particular context and often allows for interaction between the researcher and students that can add depth to the information collected. It is especially useful for studying subtleties of attitudes and behavior. Observed data, however, is not precise and cannot be generalized to larger populations. Conclusions may be suggestive rather than definitive, and others may feel that this method provides less reliable data than other collection methods.

Adapted from the California State University, Bakersfield, PACT Outcomes Assessment Handbook (1999).

Additional Resources:

Babbie, E. (1995). *The practice of social research (7th ed.)*. Belmont, CA: Wadsworth.

Palomba, C. A., & Banta, T. W. (1999). *Assessment essentials*. San Francisco: Jossey-Bass.

Oral Exams

Description: An evaluation of student knowledge levels through a face-to-face interrogative dialogue with program directors or staff responsible for maintaining statistics related to the program.

This type of tool is primarily used on students in individual programs or for a particular cohort of students. Oral exams can provide excellent results, but usually only with significant - perhaps prohibitive - additional cost. Definitely worth utilizing in programs with small numbers of students ("Low N"), and for the highest priority objectives in any program.

Strengths and Weaknesses:

Strengths

- Content and style can be geared to specific goals, objectives, and student characteristics of the institution, program, curriculum, etc.
- Specific criteria for performance can be established in relationship to curriculum
- Process of development can lead to clarification/crystallization of what is important in the process/content of student learning.
- Local grading by faculty can provide immediate feedback related to material considered meaningful.
- Greater faculty/institutional control over interpretation and use of results.
- More direct implication of results for program improvements.
- Allows measurement of student achievement in considerably greater depth and breadth through follow-up questions, probes, encouragement of detailed clarifications, etc. (= increased internal validity and formative evaluation of student abilities)
- Non-verbal (paralinguistic and visual) cues aid interpretation of student responses.
- Dialogue format decreases miscommunications and misunderstandings, in both questions and answers.
- Rapport-gaining techniques can reduce "test anxiety," helps focus and maintain maximum student attention and effort.
- Dramatically increases "formative evaluation" of student learning; i.e., clues as to how and why they reached their answers.
- Identifies and decreases error variance due to guessing.
- Provides process evaluation of student thinking and speaking skills, along with knowledge content.

Weaknesses

- Requires considerable leadership/coordination, especially during the various phases of development
- Costly in terms of time and effort (more "frontload" effort for objective; more "backload" effort for subjective)
- Demands expertise in measurement to assure validity/reliability/utility

- May not provide for externality (degree of objectivity associated with review, comparisons, etc. external to the program or institution).
- Requires considerably more faculty time, since oral exams must be conducted one-to-one, or with very small groups of students at most.
- Can be inhibiting on student responsiveness due to intimidation, face-to-face pressures, oral (versus written) mode, etc. (May have similar effects on some faculty!)
- Inconsistencies of administration and probing across students reduces standardization and generalizability of results (= potentially lower external validity).

Ways to Reduce Weaknesses

- Prearrange “standard” questions, most common follow-up probes, and how to deal with typical students’ problem responses; “pilot” training simulations.
- Take time to establish open, non-threatening atmosphere for testing.
- Electronically record oral exams for more detailed evaluation later.

Additional Resources:

Bairan, A. and B.J. Farnsworth, “*Oral Exams: An Alternative Evaluation Method*,” *Nurse Educator*, 22, Jul/Aug 1997, 6-7.

De Charruf, L.F., “*Oral Testing*,” *Mextesol Journal*, 8(2), Aug 1984, 63-79.

Dressel, J.H., “*The Formal Oral Group Exam: Challenges and Possibilities-The Oral Exam and Critical Thinking*,” Paper presented at the Annual Meeting of the National Council of Teachers of English (81st, Seattle, WA, November 22-27, 1991).

Henderson, M.L., “*Types of Classroom Tests: Essay Tests and Oral Exams*,” *American Journal of Pharmaceutical Education*, 48(3), Fall 1984, 290-292.

Nelson, J. “*Implementing Oral Exams as Part of the School Exam System*”. In: *New Approaches in the Language Classroom: Coping with Change*. Proceedings of the National Modern Languages Convention (2nd, Dublin, Ireland, January 31-February 1, 1986).

Performance Assessment

Description: Performance assessment uses student activities to assess skills and knowledge. These activities include class assignments, auditions, recitals, projects, presentations and similar tasks. At its most effective, performance assessment is linked to the curriculum and uses real samples of student work. This type of assessment generally requires students to use critical thinking and problem-solving skills within a context relevant to their field or major. The performance is rated by faculty or qualified observers and assessment data collected. The student receives feedback on the performance and evaluation.

Generally the most highly valued but costly form of student outcomes assessment - usually the most valid way to measure skill development.

Strengths and Weaknesses: Performance assessment can yield valuable insight into student learning and provides students with comprehensive information on improving their skills. Communication between faculty and students is often strengthened, and the opportunity for students' self-assessment is increased. Performance assessment, like all assessment methods, is based on clear statements about learning objectives. This type of assessment is also labor-intensive, is sometimes separate from the daily routine of faculty and student, and may be seen as an intrusion or an additional burden. Articulating the skills that will be examined and specifying the criteria for evaluation may be both time-consuming and difficult.

Advantages:

- Provide a more direct measure of what has been learned (presumably in the program)
- Go beyond paper-and-pencil tests and most other assessment methods in measuring skills
- Preferable to most other methods in measuring the application and generalization of learning to specific settings, situations, etc.
- Particularly relevant to the goals and objectives of professional training programs and disciplines with well-defined skill development.

Disadvantages:

- Ratings/grading typically more subjective than standardized tests
- Requires considerable time and effort (especially front-loading), thus being costly
- Sample of behavior observed or performance appraised may not be typical, especially because of the presence of observers

Ways to Reduce Disadvantages

- Develop specific, operational (measurable) criteria for observing and appraising performance
- Provide training for observers/appraisers
- Conduct pilot-testing in which rate of agreement (inter-rater reliability) between observers/appraisers is determined. Continue training and/or alter criteria until acceptable consistency of measurement is obtained
- Conduct observations/appraisals in the least obtrusive manner possible (e.g., use of one-way observational mirrors, videotaping, etc.)
- Observe/appraise behavior in multiple situations and settings
- Consider training and utilizing graduate students, upper level students, community volunteers, etc. as a means of reducing the cost and time demands on faculty.
- Cross-validate results with other measures, multiple methods should be used to validate the results of appraisals.

Adapted from the California State University, Bakersfield, PACT Outcomes Assessment Handbook (1999).

Additional Resources:

Angelo, T. A., & Cross, K. P. (1993). Classroom assessment techniques: A handbook for college teachers. San Francisco

Burke, Kay, ed. Authentic Assessment: A Collection. Illinois: Skylight Training and Publishing, Inc., 1992.

Hart, Diane. Authentic Assessment: A Handbook for Educators. New York: Addison-Wesley, 1994.

Jossey-Bass. Palomba, C. A., & Banta, T. W. (1999). Assessment essentials. San Francisco: Jossey-Bass.

Ryan, Alan G. "Towards Authentic Assessment in Science via STS." Bulletin of Science, Technology & Society. 1994, v 14, n 5/6, p 290

Wiggins, Grant. "The Case for Authentic Assessment." ERIC Digest. December 1990

Portfolio Evaluations

Description: Portfolios are collections of student work over time that are used to demonstrate student growth and achievement in identified areas. Portfolios can offer information about student learning, assess learning in general education and the major, and evaluate targeted areas of instruction and learning. A portfolio may contain all or some of the following: research papers, process reports, tests and exams, case studies, audiotapes, videotapes, personal essays, journals, self-evaluations and computational exercises. Portfolios are often useful and sometimes required for certification, licensure, or external accreditation reviews.

Strengths and Weaknesses: Portfolios not only demonstrate learning over time, but can be valuable resources when students apply to graduate school or for jobs. Portfolios also encourage students to take greater responsibility for their work and open lines of discussion between faculty and students and among faculty involved in the evaluation process. Portfolios are, however, costly and time-consuming and require extended effort on the part of both students and faculty. Also, because portfolios contain multiple samples of student work, they are difficult to assess and to store and may, in some contexts, require too much time and effort from students and faculty alike.

Adapted from the California State University, Bakersfield, PACT Outcomes Assessment Handbook (1999), and the University of Wisconsin, Madison, Outcomes Assessment Manual (2000).

Additional Resources:

Belanoff, P. & Belanoff, D. (1991). *Portfolios: Process and product*. Portsmouth, NH:Boynton/Cook Publishers. The Washington State University Writing Portfolio (2001).

Forrest, A. (1990). *Time will tell: Portfolio-assisted assessment of general education*. Washington, DC: AAHE Assessment Forum.

Pre-test/Post-test Evaluation

Description: This method of assessment uses locally developed and administered tests and exams at the beginning and end of a course or program in order to monitor student progression and learning across pre-defined periods of time. Results can be used to identify areas of skill deficiency and to track improvement within the assigned time frame. Tests used for assessment purposes are designed to collect data that can be used along with other institutional data to describe student achievement.

Strengths and Weaknesses: Pre-test/post-test evaluations can be an effective way to collect information on students when they enter and leave a particular program or course, and provide assessment data over a period of time. They can sample student knowledge quickly and allow comparisons between different students groups, or the same group over time. They do, however, require additional time to develop and administer and can pose problems for data collection and storage. Care should be taken to ensure that the tests measure what they are intended to measure over time (and that they fit with program learning objectives) and that there is consistency in test items, administration and application of scoring standards.

Adapted from the California State University, Bakersfield, PACT Outcomes Assessment Handbook (1999), and the University of Wisconsin, Madison, Outcomes Assessment Manual I (2000).

Additional Resources:

Berk, R. (Ed.). (1986). *Performance assessment: Methods and applications*.

Baltimore, MD. The Johns Hopkins University Press. Gronlund, N. (1991). *Measurement and evaluation in teaching (4th ed.)*. New York: MacMillan.

Palomba, C. A., & Banta, T. W. (1999). *Assessment essentials*. San Francisco: Jossey-Bass.

Reflective Essays

Description: Reflective essays may be used as an assessment tool to gauge how well students are understanding class content and issues. They are generally short essays (5 to 10 minutes) on topics related to the course curriculum and may be given as in-class assignments or homework. Reflective essays may be voluntary or required, open-ended questions on surveys required in student portfolios or capstone composition courses.

Strengths and Weaknesses: Reflective essays as an assessment tool can offer data on student opinions and perspectives at a particular moment in a class. Essays will provide a wide array of different responses and might lead to increased discussion among faculty and students. On the other hand, poorly worded, ambiguous

questions will yield little, and opinions and perceptions may vary in accuracy. Analysis of essay content also takes additional time and expertise.

Additional Resource:

Banta, T. W., Lund, J. P., Black, K. E. & Oblander, F. W. (1996). *Assessment in practice: Putting principles to work on college campuses*. San Francisco: Jossey-Bass.

Rubrics

Description: Scoring rubrics are typically grids that outline identified criteria for successfully completing an assignment or task and establish levels for meeting these criteria. Rubrics can be used to score everything from essays to performances. Holistic rubrics produce a global score for a product or performance. Primary trait analysis uses separate scoring of individual characteristics or criteria of the product or performance.

Strengths and Weaknesses: Scoring rubrics allow the instructor to efficiently and consistently look at complex products or performances and to define precise outcomes and expectations. They also are easily shared with students. Using a rubric provides several advantages to both instructors and students. Grading according to an explicit and descriptive set of criteria that is designed to reflect the weighted importance of the objective helps ensure that the grading standards don't change over time.

However, developing an effective rubric can be time-consuming and often requires ongoing edits to fine tune criteria and anticipated outcomes. Training raters to use the scoring rubrics in a consistent manner also involves a significant time commitment. Additionally, grading consistency is difficult to maintain over time because of shifting standards based on prior experience, or intrusion of other criteria.

Adapted from the California State University, Bakersfield, PACT Outcomes Assessment Handbook (1999).

Additional Resources:

Walvoord, B. E., & Anderson, V. J. (1998). *Effective grading*. San Francisco: Jossey-Bass. White, E. M. (1994). *Teaching and assessing writing*. San Francisco: Jossey-Bass.

Standardized and Local Test Instruments

Description: Selecting a standardized instrument (developed outside the institution for application to a wide group of students using national/regional norms and standards) or a locally-developed assessment tool (created within the institution, program or department for internal use only) depends on specific needs and available resources. Knowing what you want to measure is key to successful selection of standardized instruments, as is administering the assessment to a representative sample in order to develop local norms and standards. Locally-developed instruments can be tailored to measure specific performance expectations for a course or group of students.

Standardized Tests

Definition: Group administered, mostly or entirely multiple-choice, “objective” tests in one or more curricular areas. Scores are based on comparison with a reference or norm group. Typically must be obtained (purchased) from a private vender.

Target or Method: Used primarily on students in individual programs, courses or for a particular student cohort.

Advantages:

- Can be adopted and implemented quickly
- Reduce/eliminate faculty time demands in instrument development and grading (i.e., relatively low “front loading” and “back loading” effort)
- Objective scoring
- Provide for externality of measurement (i.e., external validity is the degree to which the conclusions in your study would hold for other persons in other places and at other times - ability to generalize the results beyond the original test group.)
- Provide norm reference group(s) comparison often required by mandates.
- May be beneficial or required in instances where state or national standards exist for the discipline or profession.
- Very valuable for benchmarking and cross-institutional comparison studies.

Disadvantages:

- May limit what can be measured.
- Eliminates the process of learning and clarification of goals and objectives typically associated with local development of measurement instruments.
- Unlikely to completely measure or assess the specific goals and objectives of a program, department, or institution.
- “Relative standing” results tend to be less meaningful than criterion-referenced results for program/student evaluation purposes.
- Norm-referenced data is dependent on the institutions in comparison group(s) and methods of selecting students to be tested. (Caution: unlike many norm-referenced tests such as those measuring intelligence, present norm-referenced tests in higher education do not utilize, for the most part, randomly selected or well stratified national samples.)
- Group administered multiple-choice tests always include a potentially high degree of error, largely uncorrectable by “guessing correction” formulae (which lowers validity).

- Summative data only (no formative evaluation)
- Results unlikely to have direct implications for program improvement or individual student progress
- Results highly susceptible to misinterpretation/misuse both within and outside the institution
- Someone must pay for obtaining these examinations; either the student or program.
- If used repeatedly, there is a concern that faculty may teach to the exam as is done with certain AP high school courses.

Ways to Reduce Disadvantages

- Choose test carefully, and only after staff have reviewed available instruments and determined a satisfactory degree of match between the test and the program.
- Request and review technical data, especially reliability and validity data and information on normative sample from test publishers.
- Utilize on-campus measurement experts to review reports of test results and create more customized summary reports for the institution, faculty, etc.
- Whenever possible, choose tests that also provide criterion-referenced results
- Assure that such tests are only one aspect of a multi-method approach in which no firm conclusions based on norm-referenced data are reached without cross-validation from other sources (triangulation.)
- Review curricula and coursework to assure that faculty do not teach to exam

Bottom Line:

Relatively quick, and easy, but useful mostly where group-level performance and external comparisons of results are required. Not as useful for individual student or program evaluation. May not only be ideal, but only alternative for benchmarking studies.

Locally Developed Exams

Definition: Objective and/or subjective tests designed by faculty of the program or course sequence being evaluated.

Target or Method: Used primarily on students in individual classes, a specific program of interest, or for a particular cohort of students

Advantages:

- Content and style can be geared to specific goals, objectives, and student characteristics of the program, curriculum, etc.
- Specific criteria for performance can be established in relationship to curriculum
- Process of development can lead to clarification/crystallization of what is important in the process/content of student learning.

- Local grading by faculty can provide relatively rapid feedback.
- Greater faculty/institutional control over interpretation and use of results.
- More direct implication of results for program improvements.

Disadvantages:

- Require considerable leadership/coordination, especially during the various phases of development
- Cannot be used for benchmarking, or cross-institutional comparisons.
- Costly in terms of time and effort (more “frontloaded” effort for objective; more “backloaded” effort for subjective)
- Demands expertise in measurement to assure validity/reliability/utility
- May not provide for externality (degree of objectivity associated with review, comparisons, etc. external to the program or institution).

Ways to Reduce Disadvantages:

- Enter into consortium with other programs, departments, or institutions with similar goals and objectives as a means of reducing costs associated with developing instruments. An element of externality is also added through this approach, especially if used for test grading as well as development.
- Utilize on-campus measurement experts whenever possible for test construction and validation
- Contract with faculty “consultants” to provide development and grading.
- Incorporate outside experts, community leaders, etc. into development and grading process.
- Embed in program requirements for maximum relevance with minimum disruption (e.g., a “capstone” course).
- Validate results through consensus with other data; i.e., a multi-method approach (triangulation.)

Bottom Line:

Most useful for individual coursework or program evaluation

Summary: Locally-developed instruments are directly linked to local curriculum and can identify student performance on a set of locally-important criteria. Putting together a local tool, however, is time-consuming as is development of a scoring key/method. There is also no comparison group and performance cannot be compared to state or national norms. Standardized tests are immediately available for administration and, therefore, are less expensive to develop than creating local tests from scratch. Changes in performance can be tracked and compared to norm groups and subjectivity/misinterpretation is reduced. However, standardized measures may not link to local curricula and purchasing the tests can be expensive. Test scores may also not contain enough locally-relevant information to be useful.

Adapted from the California State University, Bakersfield, PACT Outcomes Assessment Handbook (1999), and the University of Wisconsin, Madison, Outcomes Assessment Manual I (2000).

Additional Resources:

Jacobs, L. C., & Chase, C. you. (1992). Developing and using tests effectively: A guide for faculty. San Francisco: Jossey Bass.Morris,

L. L., Fitz-Gibbons, C. T., Lindheim, E. (1987). How to measure performance and use tests. Beverly Hills: Sage Publications.

National Post-Secondary Education Cooperative (NPEC) Assessment Tests Review. <http://www.nces.gov/npec/evaltests> Ory, J., & Ryan, K. E. (1993). Tips for improving testing and grading. Beverly Hills: Sage Publications.

Mazurek, D. F., “Consideration of FE Exam for Program Assessment.” Journal of Professional Issues in Engineering Education, vol. 121, no. 4, 1995, 247-249.

Scales, K., C. Owen, S. Shiohare, M. Leonard, “Preparing for Program Accreditation Review under ABET Engineering Criteria 2000: Choosing Outcome Indicators.” Journal of Engineering Education, July 1998, 207 ff.

Watson, J. L., “An Analysis of the Value of the FE Examination for the Assessment of Student Learning in Engineering and Science Topics,” Journal of Engineering Education, July 1998.

Lopez, C.L., “Assessment of Student Learning,” Liberal Education, 84(3), Summer 1998, 36-43.

Warren, J., “Cognitive Measures in Assessing Learning,” New Directions for Institutional Research, 15(3), Fall 1988, 29-39.

Student Self-efficacy

Description: Students have a sense of their own competence. Student self-efficacy involves students rating their perception of their own achievement in particular learning outcomes. Research shows a significant, although imperfect, correlation between actual and perceived competence. This assessment tool is relatively simple.

Strengths and Weaknesses: What can be problematic are gender and demographic differences in the accuracy of self-efficacy. For example, certain groups of students may rate their quantitative skills at a level below that indicated in standardized tests. Also, unless the answers are anonymous, students will be likely to over-rate their abilities. The same is true if students perceive they can be penalized by their answers.

Adapted from the Indiana University Southeast Assessment Handbook, 2004.

Student Surveys and Exit Interviews

Description: Surveys and interviews ask students to respond to a series of questions or statements about their academic experience. Questions can be both open-ended (respondents create answers) and close-ended (respondents answer from a list of simple and unambiguous responses). Surveys and interviews can be written or oral (face-to-face) or phone. Types of surveys include in-class questionnaires, mail questionnaires, telephone questionnaires, and interviews. Interviews include structured, in-person interviews and focus group interviews.

Strengths and Weaknesses: Surveys can be relatively inexpensive and easy to administer, can reach participants over a wide area, and are best suited for short and non-sensitive topics. They can give you a sense of what is happening at a given moment in time and can be used to track opinions. Data is reasonably easy to collect and tabulate, yet the sample may not be representative of the population (particularly with a low response rate). Ambiguous, poorly written items and insufficient responses may not generate enough detail for decision making. An interview can follow-up on evasive answers and explore topics in-depth, collecting rich data, new insights, and focused details. It can, however, be difficult to reach the sample and data can be time-consuming to analyze. Information may be distorted by the respondent, who may feel a lack of privacy and anonymity. The success of the interview depends ultimately on the skills of the interviewer.

Adapted from the California State University, Bakersfield, PACT Outcomes Assessment Handbook (1999), and the University of Wisconsin, Madison, Program Assessment Tool Kit (1998).

Additional Resources:

Dillman, D. (1978). *Mail and telephone surveys: The total design method*. New York: Wiley-Interscience Publication.

Fowler, F. J. (1985). *Survey research methods*. Beverly Hills: SAGE Publications.

Syllabus Analysis

Description: Syllabus analysis (as well as systematic review of textbooks, exams and other curricular material) involves looking at the current course syllabus (written or oral assignments, readings, class discussions/projects and course expectations) to determine if the course is meeting the goals and objectives that the instructor or department has set for it.

Strengths and Weaknesses: Use syllabus analysis when you want to clarify learning objectives; explore differences and similarities between sections of a course; or assess the effectiveness of instructional materials. Syllabus analysis can provide invaluable information to enhance any assessment plan. However, this review is time consuming and, as there may be more than one reviewer, there may not be adequate consistency in collecting and analyzing the data.

Additional Resources:

Bers, T., Davis, D., & Taylor, W. (1996, Nov. -Dec.). Syllabus analysis: What are

you teaching and telling your students? *Assessment Update* (8), 6, pp. 1-2, 14-15. Palombo et al. (2000). *Assessment workbook*.

Walvoord, B. E., & Anderson, V. J. (1998). *Effective grading*. San Francisco: Jossey-Bass. White, E. M. (1994). *Teaching and assessing writing*. San Francisco: Jossey-Bass.

Transcript Analysis

Description: Transcript analysis involves using data from student databases to explore course-taking or grade patterns of students. This tool can give you a picture of students at a certain point in their academic careers, show you what classes students took and in what order, and identify patterns in student grades. In sum, transcript analysis gives you a more complete picture of students' actual curricular experiences. Specific information can be drawn from transcripts to help answer research questions, and course pattern sequences can be examined to see if there is a coherence to the order of courses taken.

Strengths and Weaknesses: Transcript analysis is an unobtrusive method for data collection using an existing student database. This information can be linked to other variables such as sex or major, or used to measure outcomes. It is important to keep in mind, however, that course patterns may be influenced by other variables in students' lives that don't show up on their transcripts. Also, solutions that arise from results of the analysis may not be practical or easily implemented. It is critical to have specific questions whose answers can lead to realistic change before conducting the analysis.

Adapted from the California State University, Bakersfield, PACT Outcomes Assessment Handbook (1999), and the Ball State University, Assessment Workbook (1999).

Additional Resources:

Palomba, C. A., & Banta, T. W. (1999). *Assessment essentials*. San Francisco:

Jossey-Bass. Ratcliff, J. L. (1992). What can you learn from coursework patterns about improving undergraduate education? In J. L. Ratcliff (Vol. Ed.), *Assessment and curriculum reform: Vol. 80. New directions for higher education* (pp. 5-22). San Francisco: Jossey-Bass.

Value-Added Assessment (Pre- and Post testing)

Description: Value added assessment attempts to measure student growth at two distinct points in time. The idea is to measure the competency before and after the completion of a program. Thus, the change in competency is assumed to be the result of the program learning. The learning gain serves as an estimate of the program's contribution to the student's learning.

Strengths and Weaknesses:

Advantages

- Assessing the students when they first enter your program can establish a firm benchmark against which to measure growth or value-added.
- Pre-testing is especially helpful for measuring student knowledge, or cognitive learning, and skills, though somewhat less so for measuring values.
- Pre- and post-testing may work best when the student completes the entire program rather than stops-out and returns.
- Pre- and post-testing can be easily scored.
- Pre- and post-testing can be relatively easily analyzed using statistical procedures.

Disadvantages:

- While this approach measures learning gain, it is not always clear that all of the learning gain can be attributed to the program.
- While this approach measures learning gain, it does not assess the achievement of the program objectives. Targeted assessment tools that focus on specific objectives are a better means of assessment.
- Pre- and post-testing offers little useful information if the students know little or nothing about the subject of the program when they first enter it.
- Deciding how to develop meaningfully comparable pre- and post-assessments is difficult since the pre-test may have to be so basic that any additional learning could be seen as growth or value-added.
- If the assessment is not based on a highly structured program where the objectives are adhered to across the entire program, it may be difficult to demonstrate the causes of the value-added or to correlate the results of the post-test with the specific program.

Adapted from the Skidmore College Assessment Handbook (2011).