Pros and Cons of Tools for Doing Assessment

(Based on: Prus, Joseph and Johnson, Reid, “A Critical Review of Student Assessment Options”, in "*Assessment & Testing Myths and Realities*" edited by Trudy H. Bers and Mary L. Mittler, New Directions for Community Colleges, Number 88, Winter 1994, pp. 69-83. [Augmented by Gloria Rogers (Rose-Hulman Institute of Technology) with Engineering references by Mary Besterfield-Sacre (University of Pittsburgh)])

Information on a variety of instruments useful for doing assessment is given below.

1. *Tests*
    a. Commercial, norm-referenced, standard examinations
    b. Locally developed written examinations (objective or subjective designed by faculty);
    c. Oral examinations (evaluation of student knowledge levels through a face-to-face interrogative dialogue with program faculty).

2. *Competency-Based Methods*
    a. Performance Appraisals - systematic measurement of overt demonstration of acquired skills
    b. Simulations
    c. “Stone” courses (primarily used to approximate the results of performance appraisal, when direct demonstration of the student skill is impractical).

3. *Measures of Attitudes and Perceptions (can be self-reported or third party)*
    a. Written surveys and questionnaires (asking individuals to share their perceptions of their own or others' attitudes and behaviors including direct or mailed, signed or anonymous).
    b. Exit and other interviews (evaluating reports of subjects' attitudes and behaviors in a face-to-face interrogative dialogue).
    c. Focus groups

4. *External Examiner* (using an expert in the field from outside your program – usually from a similar program at another institution – to conduct, evaluate, or supplement the assessment of your students).

5. *Behavioral Observations* – including scoring rubrics and verbal protocol analysis (measuring the frequency, duration and topology of student actions, usually in a natural setting with non-interactive methods).

6. *Archival Records* (biographical, academic, or other file data available from college or other agencies and institutions).

7. *Portfolios* (collections of multiple work samples, usually compiled over time).

The following pages elaborate on these approaches.

Norm-Referenced, Standardized Exams

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 of 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 "*frontloading*" and "*backloading*" 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 faculty have reviewed available instruments and determined a satisfactory degree of match between the test and the curriculum.
- 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.

Bibliographic References:

1. Mazurek, D. F., “Consideration of FE Exam for Program Assessment.” *Journal of Professional Issues in Engineering Education*, vol. 121, no. 4, 1995, 247-249.
2. 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.
3. 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.

Locally Developed Exams

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

Target of 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, with careful adherence to measurement principles. Must be supplemented for *external validity*.

Bibliographic References:

1. Banta, T.W., "Questions Faculty Ask about Assessment," Paper presented at the Annual Meeting of the American Association for Higher Education (Chicago, IL, April 1989).
2. Banta, T.W. and J.A. Schneider, "Using Locally Developed Comprehensive Exams for Majors to Assess and Improve Academic Program Quality," Paper presented at the Annual Meeting of the American Educational Research Association (70th, San Francisco, CA, April 16-20, 1986).
3. Burton, E. and R.L. Linn, "Report on Linking Study--Comparability across Assessments: Lessons from the Use of Moderation Procedures in England. Project 2.4: Quantitative Models to Monitor Status and Progress of Learning and Performance", National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA, 1993
4. Lopez, C.L., "Assessment of Student Learning," *Liberal Edu*cation, 84(3), Summer 1998, 36-43.
5. Warren, J., "Cognitive Measures in Assessing Learning," *New Directions for Institutional Research*, 15(3), Fall 1988, 29-39.

Oral Examination

Definition: An evaluation of student knowledge levels through a face-to-face interrogative dialogue with program faculty.

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

Advantages

- 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.

Disadvantages

- 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 Disadvantages

- 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.

Bottom Line:

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.

Bibliographic References:

1. Bairan, A. and B.J. Farnsworth, "Oral Exams: An Alternative Evaluation Method," *Nurse Educator*, 22, Jul/Aug 1997, 6-7.
2. De Charruf, L.F., "Oral Testing," *Mextesol Journal*, 8(2), Aug 1984, 63-79.

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

Definition: A *competency-based* method whereby pre-operationalized abilities are measured in most direct, real-world approach. Systematic measurement of overt demonstration of acquired skills.

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

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.

Bottom Line:
Generally the most highly valued but costly form of student outcomes assessment – usually the most *valid* way to measure skill development.

Bibliographic References:

1. Burke, Kay, ed. *Authentic Assessment: A Collection*. Illinois: Skylight Training and Publishing, Inc., 1992.
2. Hart, Diane. *Authentic Assessment: A Handbook for Educators*. New York: Addison-Wesley, 1994.
3. Ryan, Alan G. “Towards Authentic Assessment in Science via STS.” *Bulletin of Science, Technology & Society*. 1994, v 14, n 5/6, p 290.
4. Wiggins, Grant. “The Case for Authentic Assessment.” *ERIC Digest*. December 1990.

Simulations

Definition: A *competency based* measure whereby pre-operationalized abilities are measured in most direct, real-world approach. Simulation is primarily utilized to approximate the results of performance appraisal, but when – due to the target competency involved, logistical problems, or cost – direct demonstration of the student skill is impractical.

Target of Method: Used primarily on students in individual classes or a group of students

Advantages

- Better means of evaluating depth and breadth of student skill development than tests or other performance-based measures (= *internal validity*).
- More flexible; some degree of simulation can be arranged for virtually any student target skill.
- For many skills, can be group administered, thus providing and excellent combination of quality and economy.

Disadvantages

- For difficult skills, the higher the quality of simulation the greater the likelihood of the problems of performance appraisal; e.g., cost, subjectivity, etc. (see "Performance Appraisals").
- Usually requires considerable "*frontloading*" effort; i.e., planning and preparation.
- More expensive than traditional testing options in the short run.

Ways of Reducing Disadvantages

- Reducing problems is relatively easy, since degree of simulation can be matched for maximum *validity* practicable for each situation.
- Can often be "standardized" through use of computer programs (=enhanced *external validity*).

Bottom Line:
An excellent means of increasing the *external* and *internal validity* of skills assessment at minimal long-term costs.

Bibliographic References:

1. Darling-Hammond, Linda. Jacqueline Ancess, and Beverly Falk. *Authentic Assessment in Action*. New York: Teachers College, Press, 1995.
2. Kerka, Sandra. "Techniques for Authentic Assessment." *ERIC Clearinghouse on Adult, Career, and Vocational Education*. Columbus, Ohio. 1995.
3. Paris, Scott G., and Linda R. Ayres. *Becoming Reflective Students and Teachers with Portfolios and Authentic Assessment.* Washington, DC: American Psychological Association, 1994.
4. Ryan, Alan G. "Towards Authentic Assessment in Science via STS." *Bulletin of Science, Technology & Society*. 1994, v 14, n 5/6, p 290.

"Stone" Courses[1]

[1] Often not considered an assessment method in itself.

Definition: Courses, usually required for degree/program completion, which in addition to a full complement of instructional objectives, also serve as primary vehicles of student assessment for program evaluation purposes; e.g., Capstone, Cornerstone, and Keystone courses.

Advantages:

- Provides for a synergistic combination of instructional and assessment objectives.
- A perfect mechanism for course-embedded assessment of student learning and development (i.e., outcomes, pre-program competencies and/or characteristics, "critical indicators," etc.)
- Can add impetus for design of courses to improve program orientation/integration/updating information for students.

Disadvantages:

- None specified

Ways to Reduce Disadvantages:

- None specified

Bottom Line

"Stone" course are-perfect blends of assessment and instruction to serve program quality improvement and accountability goals (capstones for outcomes measures; cornerstones for pre-program measures); and should be considered by all academic programs.

Bibliographic References:

1. Brouse, P. S., "Senior Design Project: ABET 2000 Certification, *Proceedings of the 1999 Frontiers in Education Conference*, Session 11b2-1.
2. Fong, B., "Assessment the Department Major," in Assessing Students' Learning, J. H. McMillan, ed. *New Directions in Teaching and Learning*, No. 34, San Francisco: Jossey-Bass, 1988, 71-83.
3. Michalson, W., and R. Labonte, "Capstone Design in the ECE Curriculum: Assessing the Quality of Undergraduate Projects at WPI," *1997 ASEE Annual Conference Proceedings*.
4. Shaeiwitz, J. A., "Outcomes Assessment in Engineering Education," *Journal of Engineering Education*, July 1996.
5. Trevisan, M. S., D. C. Davis, R. W. Crain, D. E. Calkins, K. L. Gentili, "Developing and Assessing Statewide Competencies for Engineering Design," *Journal of Engineering Education*, April 1998.
6. Worthen, B. R., J. R. Sanders, and J. L Fitzpatrick, *Program Evaluation: Alternative Approaches and Practical Guidelines*, New York: Longman, 1997.

Open and Closed Form Written Surveys/Questionnaires

Definition: Asking individuals to share their perceptions of their own attitudes and/or behaviors or those of others. Includes direct or mailed, signed or anonymous.

Target of Method: Used primarily on students, could be used by third parties, such as student peers, faculty, employers, parents, etc.

Advantages:

- Typically yield the perspective that students, alumni, the public, etc., have of the institution which may lead to changes especially beneficial to relationships with these groups.
- Convey a sense of importance regarding the opinions of constituent groups
- Can cover a broad range of content areas within a brief period of time
- Results ten to be more easily understood by lay persons
- Can cover areas of learning and development which might be difficult or costly to assess more directly.
- Can provide accessibility to individuals who otherwise would be difficult to include in assessment efforts (e.g., alumni, parents, employers).

When 'third-parties' are making the reports there are additional advantages, as follows:

- Can provide unique stakeholder input, valuable in its own right (especially employers and parents). How is our college serving their purposes?
- Offer different perspectives, presumably less biased than either student or assessor.
- Enable recognition and contact with important, often under-valued constituents. Relations may improve by just asking for their input.
- Can increase both *internal validity* (through "*convergent validity*"/"*triangulation*" with other data) and external validity (by adding more "natural" perspective).
- Convey a sense of importance regarding the opinions of stakeholder groups.

Disadvantages

- Results tend to be highly dependent on wording of items, *salience* of survey or questionnaire, and organization of instrument. Thus, good surveys and questionnaires are more difficult to construct than they appear.
- Frequently rely on volunteer samples which tend to be biased.
- Mail surveys tend to yield low response rates.
- Require careful organization in order to facilitate data analysis via computer for large samples.
- Commercially prepared surveys tend not to be entirely relevant to an individual institution and its students.
- Forced response choices may not allow respondents to express their true opinions.
- Results reflect *perceptions* which individuals are willing to report and thus tend to consist of indirect data.
- Locally developed instrument may not provide for *externality* of results.

Third party disadvantages also include:

- As with any indirect data, inference and reports risk high degree of error.
- Third-parties can be biased too, in directions more difficult to anticipate than self-reports.
- Less investment by third-parties in assessment processes often means lower response rates, even lower than student/alumni rates.
- Usually more logistical, time-and-motion problems (e.g., identifying sample, making contact, getting useful responses, etc.), therefore more costly than it looks.
- If information about individuals is requested, confidentiality becomes an important and sometimes problematic issue that must be addressed carefully.

Ways to Reduce Disadvantages:

- Use only carefully constructed instruments that have been reviewed by survey experts
- Include *open-ended*, respondent worded items along with *forced-choice*.
- If random sampling or surveying of the entire target population is not possible, obtain the maximum sample size possible and follow-up with nonrespondents (preferably in person or by phone).

- If commercially prepared surveys are used, add locally developed items of relevance to the institution.
- If locally developed surveys are used, attempt to include at least *some externally-referenced* items (e.g., from surveys for which national data are available).
- Word reports cautiously to reflect the fact that results represent perceptions and opinions respondents are willing to share publicly.
- Use pilot or "try out" samples in local development of instruments and request formative feedback from respondents on content clarity, sensitivity, and format.
- Cross-*validate* results through other sources of data through *triangulation*.

Ways to Reduce Third Party Disadvantages

- Very careful, explicit directions for types and perspectives of responses requested can reduce *variability*.
- Attain informed consent in cases where information about individuals is being requested.
- Coordinate contacts with other campus organs contacting the same groups, to reduce "harassment" syndrome and increase response rates.

Bottom Line:
A relatively inexpensive way to collect data on important evaluative topics from a large number of respondents. Must always be treated cautiously, however, since results only reflect what subjects are willing to report about their perception of their attitudes and/or behaviors.

Bibliographic References:

1. Converse, Jean M. & Stanley Presser (1986). *Survey Questions: Handcrafting the Standardized Questionnaire*. Sage University Paper series on Quantitative Applications in the Social Sciences, series No. 07-063. Newbury Park, CA: Sage.
2. Dovidio, John & Russell Fazio (1991). "New Technologies for the Direct and Indirect Assessment of Attitudes." In J. Tanur (ed.), *Questions About Questions: Inquires into the Cognitive Bases of Surveys*, pp. 204-237. New York: Russell Sage Foundation.
3. Sudman, Seymour & Norman Bradburn (1982). *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco: Jossey-Bass Publishers.
4. Labaw, Patricia (1981). *Advanced Questionnaire Design*, Abt Books, Incorporated.
5. Lees-Haley, Paul (1980) *Questionnaire Design Handbook*, Rubicon.
6. Fowler, Floyd J. (1993). *Survey Research Methods, 2nd Ed*. Newbury Park, CA: Sage.
7. Rossi, Peter H., James D. Wright, & Andy B. Anderson (1983*). Handbook of Survey Research*. London: Academic Press.
8. Spector, P.E. (1992). *Summated Rating Scale Construction: An Introduction*. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-082. Newbury Park, CA: Sage.
9. Suskie, Linda (1996). *Questionnaire Survey Research: What Works?* Association for Institutional Research, Resources for Institutional Research, Number Six.

Exit Interview and Other Interviews

Definition: Asking individuals to share their perceptions of their own attitudes and/or behaviors or those of others. Evaluating student reports of their attitudes and/or behaviors in a face-to-face interrogative dialogue.

Target of Method: Used primarily on students; could be used by third parties, such as student peers, employers, etc.

Advantages

- Student interviews tend to have most of the attributes of surveys and questionnaires with the exception of requiring direct contact, which may limit accessibility to certain populations. Exit interviews also provide the following additional advantages:
- Allow for more individualized questions and follow-up probes based on the responses of interviewees.
- Provide immediate feedback
- Include same observational and *formative* advantages as oral examinations.
- Frequently yield benefits beyond data collection that comes from opportunities to interact with students and other groups.
- Can include a greater variety of items than is possible on surveys and questionnaires, including those that provide more direct measures of learning and development.

When 'third-parties' are making the reports there are additional advantages, as follows:

- Can provide unique *stakeholder* input, valuable in its own right (especially employers and parents). How is the college/program/project/course serving the purposes of the *stakeholder* group?
- Offer different perspectives, presumably less biased than either student or assessor.
- Enable recognition and contact with important, often under-valued constituents. Relations may improve by just asking for their input.
- Can increase both *internal validity* (through "*convergent validity*"/"*triangulation*" with other data) and *external validity* (by adding more "natural" perspective).

Disadvantages

- Require direct contact, which may be difficult to arrange.
- May be intimidating to interviewees, thus biasing results in the positive direction.
- Results tend to be highly dependent on wording of items and the manner in which interviews are conducted.
- Time consuming, especially if large numbers of persons are to be interviewed.

Third party report disadvantages:

- As with any indirect data, inference and reports risk high degree of error.
- Third-parties can be biased too, in directions more difficult to anticipate than self-reports.
- Less investment by third-parties in assessment processes often means lower response rates, even lower than student/alumni rates.
- Usually more logistical, time-and-motion problems (e.g., identifying sample, making contact, getting useful responses, etc.), therefore more costly than it looks.
- If information about individuals is requested, confidentiality becomes an important and sometimes problematic issue that must be addressed carefully.

Ways to Reduce Disadvantages

- Plan the interviews carefully with assistance from experts
- Provide training sessions for interviewers that include guidance in putting interviewees at ease and related interview skills.
- Interview random samples of students when it is not feasible to interview all.
- Conduct telephone interviews when face-to-face contact is not feasible.
- Develop an interview format and questions with a set time limit in mind.
- Conduct pilot-testing of interview and request interviewee formative feedback.
- Interview small groups of individuals when individual interviewing is not possible or is too costly.

Ways to Reduce Third Party Disadvantages

- Conduct face-to-face or phone interviews wherever possible, increasing *validity* through probing and formative evaluation during dialogue.
- Very careful, explicit directions for types and perspectives of responses requested can reduce *variability*.
- Attain informed consent in cases where information about individuals is being requested.
- Coordinate contacts with other campus organs contacting the same groups, to reduce "harassment" syndrome and increase response rates.

Bottom Line:

Interviews provide opportunities to cover a broad range of content and to interact with respondents. Opportunities to follow-up responses can be very valuable. Direct contact may be difficult to arrange, costly, and potentially threatening to respondents unless carefully planned.

Bibliographic References:

1. Dobson, Ann (1996), *Conducting Effective Interviews: How to Find out What You Need to Know and Achieve the Right Results*, Trans-Atlantic Publications, Inc.
2. Bradburn, Norman and Seymour Sudman (?) *Improving Interview Method and Questionnaire Design*, Books on Demand (ISBN: 0835749703)

Focus Groups[2]

[2] The material for this method was developed by Gloria Rogers and colleagues at Rose-Hulman Institute of Technology

Definition: To discuss a particular topic related to a research or evaluation question with the direction of a moderator. Typically conducted with 7-12 individuals who share certain characteristics that are related to the topic of discussion. Group discussion is conducted (several times, if possible) with similar types of participants to identify trends/patterns in perceptions. Moderator's purpose is to provide direction and set the tone for the group discussion, encourage active participation from all group members, and manage time. Moderator must not allow own biases to enter, verbally or nonverbally. Careful and systematic analysis of the discussions provides information about how a product, service, or opportunity is perceived.

Target of Method: Used primarily on students, could be used by third parties, such as employers, department's visiting board, etc.

Advantages

- Useful to gather ideas, details, new insights, and to improve question design.
- Inexpensive, quick information tool, helpful in the survey design phase.
- Can aid the interpretation of results from mail or telephone surveys.
- Can be used in conjunction with quantitative studies to confirm/broaden one's understanding of an issue.
- Allows the moderator to probe and explore unanticipated issues.

Disadvantages

- Not suited for *generalizations* about population being studied.
- Not a substitute for systematic evaluation procedures.
- Moderators require training.
- Differences in the responses between/among groups can be troublesome.
- Groups are difficult to assemble.
- Researcher has less control than in individual interviews.
- Data are complex to analyze.

Ways to Reduce Disadvantages

- Offer a monetary incentive for participants if possible.
- Over-recruit participants.
- Train moderators to use open-ended questions, pauses and probes, and learn when and how to move into new topic areas.
- Have a basic understanding that focus groups are essentially an exercise in group dynamics.

Bottom Line:
Focus groups are a quick and, if locally done, inexpensive method of gathering information. They are very useful for triangulation to support other assessment methods but they are not a substitute for systematic evaluation procedures. Focus Groups should meet the same rigor as other assessment methods and should be developed and analyzed according to sound qualitative practices.

Bibliographic References:

1. Morgan, D., et. al. (1998) *Focus Groups as Qualitative Research*, University Paper series on Quantitative Applications in the Social Sciences, Newbury Park, CA: Sage.
2. Morgan, D. (1998) *Focus Groups as Qualitative Research,* Thousand Oaks, CA: Sage.
3. Krueger, Richard (1998). *Developing Questions for Focus Groups*, Vol 3. University Paper series on Quantitative Applications in the Social Sciences, Newbury Park, CA: Sage.
4. Steward, D. and P. Shamdasani (1990). *Focus Groups: Theory and Practice*, University Paper series on Quantitative Applications in the Social Sciences, Newbury Park, CA: Sage.
5. Krueger, Richard (1997). *Moderating Focus Groups*, Vol 4. University Paper series on Quantitative Applications in the Social Sciences, Newbury Park, CA: Sage.
6. Morgan, D., and A. Scannell (1997). *Planning Focus Groups*, Vol 2. University Paper series on Quantitative Applications in the Social Sciences, Newbury Park, CA: Sage.

External Examiner

Definition: Using an expert in the field from outside your program, usually from a similar program at another institution to conduct, evaluate, or supplement assessment of your students. Information can be obtained from external evaluators using many methods including surveys, interviews, etc.

Target of method: Used primarily on students in individual classes or for a particular cohort of students; could be used by third parties, such as employers or visiting board, etc.

Advantages:

- Increases impartiality, third party objectivity (=*external validity*)
- Feedback useful for both student and program evaluation. With a knowledgeable and cooperative (or well-paid) examiner, provides an opportunity for a valuable program consultation.
- May serve to stimulate other collaborative efforts between departments/institutions - Incorporate external *stakeholders* and communities
- Students may disclose to an outsider what they might not otherwise share
- Outsiders can "see" attributes to which insiders have grown accustomed
- Evaluators may have skills, knowledge, or resources not otherwise available
- Useful in conducting *goal-free evaluation* (discovery-based evaluation without prior expectations)

Disadvantages:

- Always some risk of a misfit between examiner's expertise and/or expectations and program outcomes
- For individualized evaluations and/or large programs, can be very costly and time consuming
- Volunteers may become "donor weary"

Way to Reduce Disadvantages:

- Share program philosophy and objectives – and agree on assessment criteria - beforehand.
- Form reciprocal external examiner "consortia" among similar programs to minimize costs, swapping external evaluations back and forth.
- Limit external examiner process to program areas where *externality* may be most helpful.

Bottom Line:
Best used as a supplement to your own assessment methods to enhance external validity, but not as the primary assessment option. Other benefits can be accrued from the cross-fertilization that often results from using external examiners.

Bibliographic References:

1. Bossert, James L., *Quality Function Deployment*, Milwaukee: ASQC Quality Press, 1991, especially pp. 52-64.
2. Fitzpatrick, Jody L. and Michael Morris, Eds*., Current and Emerging Ethical Challenges in Evaluation*, San Francisco, CA: Jossey-Bass, 1999.

Behavioral Observations

Definition: Measuring the frequency, duration, *topology*, etc. of student actions, usually in a natural setting with non-interactive methods. For example, formal or informal observations of a classroom. Observations are most often made by an individual and can be augmented by audio or videotape.

Target of Method: Used primarily on individuals or groups of students in classes

Advantages

- Best way to evaluate degree to which attitudes, values, etc. are really put into action (= most *internal validity*).
- Catching students being themselves is the most "natural" form of assessment (= best *external validity*).
- Least intrusive assessment option, since purpose is to avoid any interference with typical student activities.

Disadvantages

- Always some risk of *confounded* results due to "*observer effect*;" i.e., subjects may behave atypically if they know they're being observed.
- Depending on the target behavior, there may be socially or professionally sensitive issues to be dealt with (e.g., invasion of privacy on student political activities or living arrangements) or even legal considerations (e.g., substance abuse or campus crime).
- May encourage "Big Brother" perception of assessment and/or institution.
- Inexperienced or inefficient observers can produce unreliable, invalid results.

Ways to Reduce Disadvantages

- Avoid socially or ethically sensitive target behaviors, especially initially.
- Include representative student input in process of determining "sensitivity" of potential target behaviors.
- Utilize electronic "observers: (i.e., audio and video recorders) wherever possible, for highly accurate, reliable, permanent observation record (although this may increase assessment cost in the short run if equipment is not already available.)
- Strictly adhere to ethical guidelines for the protection of human research subjects.

Bottom Line:
This is the best way to know what students actually do, how they manifest their motives, attitudes and values. Special care and planning are required for sensitive target behaviors, but it's usually worth it for highly *valid*, useful results.

Bibliographic References:

1. Lincoln, Y. S. and E. G. Guba (1985). *Naturalistic Inquiry*. Newbury Park, CA, SAGE Publications.
2. Miles, M. B. and A. M. Huberman (1984). *Qualitative Data Analysis*. Beverly Hills, Sage Publications.

Archival Data

Definition: Biographical, academic, or other file data available from college or other agencies and institutions.

Target of Method: Primarily aggregated student information; can use comparable data from other institutions for benchmarking.

Advantages:

- Tend to be accessible, thus requiring less additional effort.
- Build upon efforts that have already occurred.
- Can be cost efficient if required date 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") – is there someone on campus responsible for "Institutional Research."
- 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.

Bottom Line:
Can be quick, easy, and cost-effective method, if data is available and accessible. Usually limited data quality but integral to valuable longitudinal comparisons. Should be a standard component of all assessment programs.

Bibliographic References:

1. Astin, Alexander W. "Involvement in Learning Revisted: Lessons We Have Learned." *Journal of College Student Development*; v37 n2 p. 123-34, March 1996.
2. 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.

Portfolios

Definition: Collections of multiple student work samples usually compiled over time. Rated by some type of *rubric*.

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

Advantages:

- Can be used to view learning and development *longitudinally* (e.g. samples of student writing over time can be collected), which is most valid and useful perspective.
- Multiple components of a curriculum can be measured (e.g., writing, critical thinking, research skills) at the same time.
- Samples in a portfolio are more likely than test results to reflect student ability when pre-planning, input from others, and similar opportunities common to most work settings are available (which increases *generalizability/external validity* of results).
- The process of reviewing and grading portfolios provides an excellent opportunity for faculty exchange and development, discussion of curriculum goals and objectives, review of grading criteria, and program feedback.
- Economical in terms of student time and effort, since no separate “assessment administration” time is required.
- Greater faculty control over interpretation and use of results.
- Results are more likely to be meaningful at all levels (i.e., the individual student, program, or institution) and can be used for diagnostic/prescriptive purposes as well.
- Avoids or minimizes “test anxiety” and other “one shot” measurement problems.
- Increases “power” of maximum performance measures over more artificial or restrictive “speed” measures on test or in-class sample.
- Increases student participation (e.g., selection, revision, evaluation) in the assessment process.

Disadvantages

- Costly in terms of evaluator time and effort.
- Management of the collection and grading process, including the establishment of reliable and valid grading criteria, is likely to be challenging.
- May not provide for *externality*.
- If samples to be included have been previously submitted for course grades, faculty may be concerned that a hidden agenda of the process is to validate their grading.
- Security concerns may arise as to whether submitted samples are the students’ own work, or adhere to other measurement criteria.

Ways to Reduce Disadvantages

- Consider having portfolios submitted as part of a course requirement, especially a “capstone course” at the end of a program.
- Utilize portfolios from representative samples of students rather than having all students participate (this approach may save considerable time, effort, and expense but be problematic in other ways).
- Have more than one rater for each portfolio; establish *inter-rater reliability* through piloting designed to fine-tune rating criteria.
- Provide training for raters.
- Recognize that portfolios in which samples are selected by the students are likely represent their best work.
- Cross-validate portfolio products with more controlled student work samples (e.g., in-class tests and reports) for increased *validity* and security.

Bottom Line:
Portfolios are a potentially valuable option adding important longitudinal and “qualitative” data, in a more natural way. Particular care must be taken to maintain *validity*. Especially good for multiple-objective assessment.

Bibliographic References:

1. Barrett, H.C. (1994). *Technology-supported assessment portfolios.* "Computing Teacher," 21(6), 9-12. (EJ 479 843)

2. Hart, D. (1994). *Authentic assessment: a handbook for educators.* Menlo Park, CA: Addison-Wesley.
3. Hodges, D. (1998). *Portfolio: A self-learning guide.* Barrington, IL.
4. Jackson, L. and Caffarella, R.S. (1994). *Experiential learning: A new approach.* San Francisco, CA: Jossey-Bass.
5. Khattru, N., Kane, M., and Reeve, A. (1995). *How performance assessments affect teaching and learning.* Educational Leadership. (11), 80-83.
6. Murphy, S.M. (1998). Reflection: In portfolios and beyond. *Clearing House,*(72), 7-10.
7. Paulson, L.F., Paulson, P.R., & Meyer, C. (1991) *What makes a portfolio a portfolio?* "Educational Leadership," 48(5), 60-63. (EJ 421 352)
8. Porter, C. and Cleland, J. (1995). *The portfolio as a learning strategy.* Portsmouth, NH: Boynton/Cook Publishers.
9. Rogers, Gloria and Timothy Chow, “Electronic Portfolios and the Assessment of Student Learning.” *Assessment Update*, Josey-Bass Publisher, January-February 2000, Vol. 12, No. 1, pp. 4-6, 11.