

AABInternational

ASSESSMENT GUIDE

Form 223

JANUARY 2008

AVIATION ACCREDITATION BOARD INTERNATIONAL

3410 SKYWAY DRIVE
AUBURN, ALABAMA 36830

INTRODUCTION

The purpose of this Guide is to provide aviation departments a resource to understanding the terminology used throughout all stages of the AABI accreditation process.

While AABI recognizes and supports the prerogative of institutions to use and adopt the terminology of their choice, it is necessary for all to have a consistent understanding of terminology. AABI also recognizes that there may be some minor differences in terminology; however, it is important that the fundamental interpretations of the major terms be consistent. With that purpose in mind, the AABI will use the following basic definitions (boldface) and explanatory text (italics).

Program Educational Objectives – Program educational objectives are broad statements that describe the career and professional accomplishments that the program is preparing graduates to achieve.

These are also often referred to by institutions as goals, career outcomes, or standards. There are two types of objectives – those that all graduates are expected to accomplish and those that some subgroups, but not all graduates, are expected to accomplish. The audiences for objective statements are normally external constituents, such as prospective students, employers, and transfer institutions.

General Outcomes – General outcomes are narrower statements that describe what students are expected to know and able to do by the time of graduation. These relate to the skills, knowledge, and behaviors that students acquire in their matriculation through the program.

These are also often referred to by institutions as graduation objectives, goals, or standards. Achievement of all of the general outcomes should indicate that the graduate is equipped to achieve the corresponding program education objectives.

Assessment – Assessment is one or more processes that identify, collect, and prepare data to evaluate the achievement of general outcomes and program educational objectives.

Often, the entire process is referred to as assessment, and the program or institution does not subdivide the overall process into component parts. While assessment data are useful for display during accreditation reviews, data alone do not provide the documented evidence for continuous improvement.

Evaluation – Evaluation is one or more processes for interpreting the data and evidence accumulated through assessment practices. Evaluation determines the extent to which general outcomes or program educational objectives are being achieved, and results in decisions and actions to improve the program.

Overview of Assessment Methods*

1. Written surveys and questionnaires (asking individuals to share their perceptions about the study target – e.g. their own or others’ skills/attitudes/behavior, or program/course qualities and attributes). (Page 5)
2. Exit and other interviews (asking individuals to share their perceptions about the target of study – e.g. their own skills/attitudes, skills and attitudes of others, or program qualities—in a face-to-face dialog with an interviewer). (Page 7)
3. Commercial, non-referenced, standardized examinations (commercially developed examinations, generally group administered, mostly multiple choice, “objective” tests, usually purchased from a private vendor). (Page 10)
4. Locally developed examinations (objective or subjective designed by local staff/faculty). (Page 12)
5. Archival records (biographical, academic, or other file data available from college or other agencies and institutions). (Page 14)
6. Focus groups (guided discussion of a group of people who share certain characteristics related to the research or evaluation question, conducted by trained moderator). (Page 16)
7. Portfolios (collections of work samples, usually compiled over time and rated using rubrics). (Page 18)
8. Simulations (a competency based measure where a person’s abilities are measured in a situation that approximates a “real world” setting). Simulation is primarily used when it is impractical to observe a person performing a task in a real world situation (e.g., on the job). (Page 20)
9. Performance appraisals (systematic measurement of overt demonstration of acquired skills, generally through direct observation in a “real world” situation – e.g. while student is working on internship or on project for client).
10. 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). (Page 21)
11. Oral examinations (evaluation of student knowledge levels through a face-to-face dialogue between the student and the examiner, usually faculty). (Page 22)

12. Behavioral observations (measuring the frequency, duration and context of subject's actions, usually in a natural setting with non-interactive methods). (Page 24)

13. Core Values and Glossary (Page 25)

*Except where noted, materials relating to the pros and cons of assessment methods have been modified by Gloria Rogers and used with permission. Pres, J. and Johnson, R., "Assessment & Testing Myths and Realities." *New directions for Community Colleges*, No. 88, Winter 94.

Written Surveys/Questionnaires

Definition: Asking individuals to share their perceptions about the study target—e.g. their own or others skills/attitudes/behavior, or program/course qualities and attributes.

Advantages:

- Typically yield the perspective that students, alumni, the public, etc., have of the institution that may lead to changes especially beneficial to relationships with these groups.
- Can cover a broad range of attributes within a brief period of time.
- Results tend to be more easily understood by laypersons.
- Can cover areas of developments, 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 completing the survey/questionnaire there are additional advantages, as follows:

- Can provide unique **stakeholder** input, valuable in its own right (especially employers and parents). How is the course/program/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**.
- 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 can 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 provide opportunities for 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 include:

- As with any indirect data, inference and reports can contain a 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 requires logistical details (e.g., identifying sample, making contact, getting useful responses, etc.), therefore more costly than it looks.
- If information about specific 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 non-respondents (preferably in person or by phone).

Exit 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 dialogue.

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 provide the following advantages:

- Allow for more individualized questions and follow-up probes based on the responses of interviewees.
- Provide immediate feedback to interviewer.
- 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 the 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:

- Requires 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 working 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.

- Usually requires logistical details (e.g., identifying sample, making contact, getting useful responses, etc.), therefore more costly than it looks.
- If information about specific 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 purposeful 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 feedback from interviewee to improve the interview process.
- Utilize focus groups 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 during dialogue.
- 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 of responses requested can reduce variability.
- Attain informed consent in cases where information about specific individuals is being requested.
- Coordinate contacts with other campus organizations 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:

Sudman, Seymour & Norman Bradburn (1982). *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco: Jossey-Bass Publishers.

Suskie, Linda (1996). *Questionnaire Survey Research: What Works?* Association for Institutional Research, Resources for Institutional Research, Number Six.

Commercial, 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 purchased from a private vendor.

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., *externality 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* group(s) comparison often required by mandates.
- May be beneficial or required in instance 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* (define & insert) 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*).
- Results unlikely to have direct implications for program improvement or individual student progress.
- Results highly susceptible to misinterpretations/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 the test carefully, and only after faculty have reviewed available instruments and determined a satisfactory degree of match between the test and the learning outcomes of 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 *validation* from other sources (*triangulation*).
- 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 organizations 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:

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

Locally Developed Exams

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

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 exams; more “*backloaded*” effort for subjective exams).
- Demands expertise in measurement to assure *validity/reliability/utility*.
- May not provide for *externality*.

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.
- 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 use of 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 Reference:

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).

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 Education a Research Association (70th, San Francisco, CA, April 16-20, 1986).

Burton, E. and R.L. Linn, "Report on Linking Study—Comparability across Assessment: Lessons Status and Progress of Learning and Performance", National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA, 1993

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.

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 many times the only alternative for benchmarking studies.

Bibliographic References:

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.

Archival Records

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

Advantages:

- Tend to be accessible, thus requiring minimal effort.
- Build upon data collection efforts that have already occurred.
- Can be cost efficient if required data is readily retrievable in desired format.
- Constitute non-intrusive measurement, not requiring additional time or effort from students or other groups.
- Very useful for longitudinal studies.
- Good 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 is 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 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; where applicable, Institutional Review Board approval should be obtained if there is doubt.
- Availability of data may discourage the development of other, more appropriate measures or data sources.
- May encourage attempts to “find ways to use data” rather than assessment 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”). An Office of Institutional Research is found on many campuses and can be helpful in this process.
- 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 are available and accessible.
Usually limited data quality but integral to valuable longitudinal comparisons. Should be a standard component of all assessment programs.

References:

Astin, Alexander W. "Involvement in Learning Revisited: Lessons We Have Learned." *Journal of College Student Development*; v37 n2 p123-34 Mar 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

Focus Groups**

Definition: Typically conducted with 7-12 individuals who share certain characteristics that are related to a particular topic related to a research or evaluation question. Group discussions are conducted by a trained moderator with participants (several times, if possible) 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 that can be used to evaluate and/or improve the desired outcome.

Advantages:

- Useful to gather ideas, details, new insights and to improve question design.
- Helpful in the design of surveys.
- Can be used to get more in-depth information on issues identified by a survey.
- Can inform 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.
- Interaction among focus group participants often leads to new insights.
- 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 can be difficult to assemble.
- Moderator has less control than in individual interviews.
- Data are complex to analyze.

Ways to Reduce Disadvantages:

- Offer an 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.

Example of Applications:

- Focus groups can be held to provide in-depth information of interest generated from a survey.
Focus groups can be used as a follow-up to survey data. In cases where the results of a survey do not meet the expected standard on a particular outcome, a focus group of participants who are representative of the population surveyed (e.g., students,

alumni, females) could be held to further investigate the results. For example, if the analysis of the questionnaires of senior students indicates that they, generally, did not feel they had adequate communication skills, a focus group of senior students could be established to examine why they believe students responded that way and what they would suggest to improve the development of communications skills.

- Focus groups can be used to get input from alumni or business partners on the strengths and weaknesses in the knowledge and/or skills of graduates. Focus groups are a particularly helpful tool to use to “triangulate” or validate the results from other assessment methods.

Examples Instruments/methodologies:

- Detailed procedure for developing focus groups can be found in Steward, D. and Shamdasani, P. (1990) cited below.]

Bottom Line:

Focus groups are a quick and, if locally done, inexpensive method of gathering information. Someone who has training and experience in conducting Focus Groups and analysis of Focus Group data should conduct them. 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:

Morgan, D., et. Al. (1998) *Focus Groups as Qualitative Research*, University Paper series on Quantitative Applications in the social Sciences, Newbury Park, CA: Sage.

Morgan, D. (1998) *Focus Groups as Qualitative Research*, Thousand Oaks, CA: Sage.

Krueger, Reichard (1998) *Developing Questions for Focus Groups*, Vol 3. University Paper Series on Quantitative Applications in the Social Sciences, Newbury Park, CA: Sage.

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.

**Prepared by Gloria Rogers, Rose-Hulman Institute of Technology

Portfolios

Definition: collections of multiple student work samples usually compiled over time and rated using *rubrics*. The design of a portfolio is dependent upon how the scoring results are going to be used.

Advantages:

- Can be used to view learning and development *longitudinally* (e.g. samples of student writing over time can be collected), which is a useful perspective.
- Multiple components of a curriculum scan are 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 preplanning, 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 scoring portfolios provides an excellent opportunity for faculty exchange and development, discussion of curriculum goals and objectives, review of scoring 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:

- Can be costly in terms of evaluator time and effort.
- Management of the collection and scoring process, including the establishment of reliable and valid scoring rubrics, 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.
- Investigate the use of electronic portfolios as a means to increase process efficiency.

- Utilize portfolios from representative samples of students rather than having all students participate (this approach may save considerable time, effort, and expense but may 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 student are likely representing 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 potentially valuable options 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:

- Barrett, H.C. (1994). Technology-supported assessment portfolios. “Computing Teacher,” 21 (6), 9-12. (EJ 479 843)
- Hart, D. (1994). Authentic Assessment: a Handbook for Educators. Menlo Park, CA: Addison-Wesley.
- Hodges, D. (1998). Portfolio: A Self-Learning Guide. Barrington, IL.
- Paulson, L.F., Paulson, P.R., & Meyer, C. (1991) What Makes a Portfolio a Portfolio?

Simulations

Definition: A competency based measure where a person's abilities are measured in a situation that approximates a "real world" setting. Simulation is primarily used when it is impractical to observe a person performing a task in a real world situation (e.g., on the job).

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 an excellent combination of quality and economy.

Disadvantages:

- For difficult skills, the higher the quality of simulation the greater the likelihood that it will suffer from same problems as "Performance Appraisals" listed below:
 - Ratings of student performance are typically more subjective than standardized tests.
 - Sample of behavior observed or performance appraised may not be typical, especially because of the presence of others.
 - 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 (and enhance external *validity*).

Bottom Line:

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

Bibliographic References:

- Darling-Hammond, Linda. Jacqueline Ancess, and Beverly Falk. *Authentic Assessment in Action*. New York: Teachers college, Press, 1995.
- Ryan, Alan G. "Towards Authentic Assessment in Science via STS." *Bulletin of Science, Technology & Society*. 1994, v 14, n 5/6, p 290.

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.

Advantages:

- Increases impartially, third party objectivity (external validity).
- Feedback useful for both student and program evaluation. With a knowledgeable examiner it 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 (without prior expectations).

Disadvantages:

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

Ways to Reduce Disadvantages:

- Share program philosophy and objectives and agree on assessment criteria before the assessment.
- 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:

Fitzpatrick, Jody L. and Michael Morris, Eds., *Current and Emerging Ethical Challenges in Evaluation*, San Francisco, CA: Jossey-Bass, 1999.

Oral Examination

(This method may be inconsistent with campus policies that prohibit the use of oral examinations.)

Definition: An assessment of student knowledge levels through a face-to-face dialogue between the student and examiner, usually faculty.

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 course/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 considerable 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.
- Can be difficult to document by note-taking and providing student feedback with a grade.
- 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 reduce 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, and for the highest priority objectives in any program and local testing policies do not prohibit the testing method.

Bibliographic References:

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

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.

Advantages:

- Best way to evaluate degree to which attitudes, values, etc. is really put into action.
- Catching students being themselves is the most “natural” form of assessment.
- 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 purpose of the data gathering, there may be socially or professionally sensitive issues to be dealt with (e.g., invasion of privacy pm 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 using this method when studying socially or ethically sensitive issues.
- Include representative student input in process of determining “sensitivity” of issue.
- Utilize electronic “observers” (i.e., audio and video recorders) wherever possible, for highly accurate, reliable, permanent observations 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 when studying sensitive issues, but in situations where the information derived from the observations is critical, it’s usually worth it for highly *valid*, useful results.

Bibliographic References:

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

Core Value 1.

Responsible membership in AABI is based on the conviction that the results of any aviation program of study must be verifiable. All academic disciplines have established canons of evidence, which they use to assess the adequacy of their scholarly products. More importantly, all members of the aviation educational community accept the proposition that it is not proper to assert conclusions without recourse to evidence.

Core Value 2.

Accreditation must constitute more than a periodic event and must lead to significant levels of ongoing program improvement. The process of accreditation, moreover, should result in more than an external validation of "quality;" it should "add value" to an institution by providing an important opportunity to inquire deeply into student learning - a matter related directly to the mission of every aviation program.

GLOSSARY

convergent validity

a general agreement among ratings, gathered independently of one another, where measures should be theoretically related.

criterion-referenced

criterion-referenced tests determine what test takers can do and what they know, not how they compare to others. Criterion-referenced tests report how well students are doing relative to a pre-determined performance level on a specified set of educational goals or outcomes included in the curriculum.

externality

externality refers to the extent to which the results of the assessment can be generalized to a similar context.

external validity

external validity refers to the extent to which the results of a study is generalizable or transferable to other settings. Generalizability is the extent to which assessment findings and conclusions from a study conducted on a sample population can be applied to the population at large. Transferability is the ability to apply the findings in one context to another similar context.

forced-choice

the respondent only has a choice among given responses (e.g., very poor, poor, fair, good, very good).

formative assessment

intended to assess ongoing program/project activity and provide information to improve the project. Assessment feedback is short term in duration.

frontload (--ed, --ing)

amount of effort required in the early stage of assessment method development or data collection.

generalization (generalizability)

the extent to which assessment findings and conclusions from a study conducted on a sample population can be applied to the population at large.

goal-free evaluation

focuses on actual outcomes rather than intended general outcomes. Evaluation is done without prior knowledge of the goals of the program.

inter-rater reliability:

the degree to which different raters/observers give consistent estimates of the same phenomenon.

internal validity

internal validity refers to (1) the rigor with which the study was conducted (e.g., the study's design, the care taken to conduct measurements, and decisions concerning what was and wasn't measured) and (2) the extent to which the designers of a study have taken into account alternative explanations for any causal relationships they explore.

longitudinal studies

data collected from the same population at different points in time.

norm (--ative)

a set standard of development or achievement usually derived from the average or median achievement of a large group.

norm-reference

a norm-referenced test is designed to highlight achievement differences between and among students to produce a dependable rank order of students across a continuum of achievement from high achievers to low achievers.

observer effect

the degree to which the assessment results are affected by the presence of an observer.

open-ended

assessment questions that are designed to permit spontaneous and unguided responses.

operational (--ize)

defining a term or object so that it can be measured. Generally states the operations or procedures used that distinguish it from others.

reliability

the extent to which an experiment, test, or any measuring procedure yields the same result on repeated trials.

rubrics

a set of categories that define and describe the important components of the work being completed, critiqued, or assessed. Each category contains a graduation of levels of completion or competence with a score assigned to each level and a clear description of what criteria need to be met to attain the score at each level.

salience

as striking point or feature.

stakeholder

anyone who has a vested interest in the outcome of the program/project.

summative assessment

assessment that is done at the conclusion of a course or some larger instructional period (e.g., at the end of the program). The purpose is to determine success or to what extent the program/ project/course met its goals.

third party: person(s)

other than those directly involved in the educational process (e.g., employers, parents, and consultants).

triangulate (triangulation)

the use of a combination of assessment methods in a study. An example of triangulation would be an assessment that incorporated surveys, interviews, and observations.

topology

mapping of the relationships among subjects.

utility

usefulness of assessment results.

variable (variability)

observable characteristics that vary among individual's responses.

validity

the degree to which a study accurately reflects or assesses the specific concept that the researcher is attempting to measure. Validity has three components:

- Relevance – the option measures your educational objective as directly as possible
- Accuracy – the option measures your educational objective as precisely as possible
- Utility – the option provides formative and summative results with clear implications for education program evaluation and improvement.