

Assessment Methods*

1. **Written surveys and questionnaires** - Asking individuals to share their perceptions about a particular area of interest—e.g., their own or others' skills/attitudes/behavior, or program/course qualities and attributes.
2. **Exit and other interviews** - Asking individuals to share their perceptions about a particular area of interest—e.g., their own skills/attitudes, skills and attitudes of others, or program qualities—in a face-to-face dialog with an interviewer.
3. **Commercial, norm-referenced, standardized examinations** - Commercially developed examinations, generally group administered, mostly multiple choice, “objective” tests, usually purchased from a private vendor.
4. **Locally developed assessments** - Objective or subjective designed by local staff/faculty.
5. **Focus groups** - Guided discussion of a group of people who share certain characteristics related to the research or evaluation question, conducted by trained moderator.
6. **Portfolios** (collections of work samples, usually compiled over time and rated using scoring rubrics).
7. **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.
8. **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 students.
9. **Oral examinations** - Evaluation of student knowledge levels through a face-to-face dialogue between the student and the examiner—usually faculty.

*Except where noted, materials relating to the advantages and disadvantages of assessment methods have been modified by Gloria Rogers and used with permission. Prus, J. and Johnson, R., "Assessment & Testing Myths and Realities." New Directions for Community Colleges, No. 88, Winter 94. These materials cannot be duplicated without the expressed written consent of the authors.



GLOSSARY*

Backload (--ed, --ing): amount of effort required after the data collection.

Competency: level at which performance is acceptable.

Confounded: confused.

Convergent validity: 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 are 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: Goal-free evaluation focuses on actual outcomes rather than intended program 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: Reliability is the extent to which an experiment, test, or any measuring procedure yields the same result on repeated trials

Rubrics: A rubric is a set of categories that define and describe the important components of the work being completed, critiqued, or assessed. Each category contains a gradation 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: a 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, 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 individuals responses.

Validity: Validity refers to 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 educational program evaluation and improvement

Written Surveys/Questionnaires

Definition: Asking individuals to share their perceptions about the curricular/co-curricular areas of interest—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 program that may lead to changes especially beneficial to improving the program.
- Can cover a broad range of areas of interest within a brief period of time.
- Results tend to be more easily understood by lay persons.
- Can cover areas of interest, 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 alumni).
How is the program serving their purposes?
- Offer different perspectives, presumably less biased than either student or faculty.
- 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 (**forced-choice**) 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 also include:

- As with any indirect data, inference and reports can contain a high degree of interpretation 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 appears.
- 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).
- If commercially prepared surveys are used, add locally developed items of relevance to the program.
- 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 Lines:

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.

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/questions 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/constituent input, valuable in its own right (especially employers and alumni). How is the program/course serving the purposes of the stakeholder group?
- Offer different perspectives, presumably less biased than either student or the faculty.
- 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 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 in interpretation.
- 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 appears.
- 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 questions and process 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.
- 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 Lines:

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.

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 outside the program/ institution (e.g., accreditation agency, state or federal regulations).
- 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 is 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 objectives and outcomes of a program, department, or institution.
- “Relative standing” (i.e., how student performance compares with others) results tend to be less meaningful than **criterion-referenced** (i.e., what students know or can do without comparison to others) 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 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 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/program, 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**).

Bottom Lines:

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.

Locally Developed Assessments

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

Advantages:

- Content and style can be geared to specific outcomes, objectives, and student characteristics of the program, curriculum, etc.
- Specific indicators 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 scoring by program 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:

- When using to assess institutional/program outcomes, requires considerable leadership/coordination, especially during the various phases of development to get consistent and reliable results.
- Cannot be used for benchmarking, or cross-institutional comparisons.
- Will not provide for **externality**.

Ways to Reduce Disadvantages:

- Create teams of faculty who all deliver content related to the learning outcomes who develop a common rubric or benchmark for assessing student performance.
- Include **formative assessment** using common scoring tool to observe progress overtime and modify scoring tool when necessary.
- Utilize on-campus assessment resources whenever possible for construction of assessments and scoring tools.
- Validate results through use of multi-method approach (**triangulation**).

Bottom Lines:

Most useful for individual coursework or program evaluation, with careful adherence to assessment principles.



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 coding and 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 code and 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 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.
- 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.

Bottom Lines:

Focus groups are a quick and, if locally done, inexpensive method of gathering information. They should be conducted by someone who has training and experience in conducting Focus Groups and analysis of Focus Group data. 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.

*Prepared by Gloria Rogers, ABET, Inc.

Portfolios

Definition: Collections of multiple student work samples usually compiled over time and scored 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 can be assessed (e.g., writing, critical thinking, research skills) at the same time.
- The process of reviewing and scoring portfolios provides an excellent opportunity for faculty exchange and development, discussion of curriculum objectives and outcomes, review of scoring criteria, and program feedback.
- 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” assessments.
- 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 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.
- Educate the raters about the process.
- 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 Lines:

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-learning outcomes assessment.

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

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 assessing skills.
- Preferable to most other methods in measuring the application and **generalization** of learning to specific settings, situations, etc.
- Particularly relevant to the objectives and outcomes of professional training programs and disciplines with well defined skill development.

Disadvantages:

- Rating of student performance is 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) indicators 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 performance indicators for more specificity until acceptable consistency of measurement is obtained.
- Conduct observations/appraisals in the least intrusive manner possible
- Observe/appraise behavior in multiple situations and settings.
- Consider training and utilizing graduate students, upper level students, 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 Lines:

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

External Examiner

Definition: Using an expert in the field from outside your program such as someone from a similar program at another institution or a capstone project client to evaluate, or supplement assessment of your students. Information can be obtained from external evaluators using many methods including feedback forms (including scoring rubrics), surveys, interviews, etc.

Advantages:

- Increases impartiality, **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 business partners or other programs.
- Incorporate the use of external **stakeholders**.
- 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 program outcomes.
- For individualized evaluations and/or large programs, can be very costly and time consuming.
- Volunteers may become “donor weary” (tired from being asked multiple times to participate).

Way to Reduce Disadvantages:

- Share program philosophy and outcomes and agree on assessment procedure 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 Lines:

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.



(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 objectives and outcomes, and student characteristics of the institution, program, curriculum, etc.
- Specific indicators 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 scoring 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 knowledge 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.
- 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).
- 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, at most, with very small groups of students.
- 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 Lines:

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.