American Association for Higher Education

Principles of Good Practice for Assessing Student Learning

➢ The assessment of student learning begins with educational values.

➢ Assessment is most effective when it reflects an understanding of learning as multidimensional, integrated, and revealed in performance over time.

➢ Assessment works best when the programs it seeks to improve have clear, explicitly stated purposes.

➢ Assessment requires attention to outcomes but also and equally to the experiences that lead to those outcomes.

➢ Assessment works best when it is ongoing, not episodic.

➢ Assessment fosters wider improvement when representatives from across the educational community are involved.

➢ Assessment makes a difference when it begins with issues of use and illuminates questions that people really care about.

➢ Assessment is most likely to lead to improvement when it is part of a larger set of conditions that promote change.

➢ Through assessment, educators meet responsibilities to students and to the public.

TABLE OF CONTENTS

INTRODUCTION

This manual is intended to assist Norfolk State University faculty members in developing assessment plans at the departmental and program levels by providing a brief introduction to outcomes assessment processes. This document is compiled from various sources (see *Bibliography*) and reflects current best practices in assessment of student learning outcomes.

Institutional Effectiveness and Assessment staff will provide further consultation and support to assist in the development of meaningful, manageable, and sustainable assessment practices. By invitation, the staff will help faculty and administrators:

- develop mission, goals, and outcomes statements,
- select appropriate assessment methods, ANF
- develop and administer assessment procedures and analyze their results.

Definition

Assessment is the systematic collection, review, and use of information about educational programs undertaken for improving student learning and development. The goal of assessment is to examine the quantitative and qualitative evidence generated about student competence, to use this evidence to improve the learning of current and future students, and to present this information to external stakeholders.

In other words, assessment is the process of gathering, analyzing and discussing information from multiple and diverse sources in order to develop a deep understanding of what students know, understand, value, and can do with their knowledge and skills as a result of their experiences; the process culminates when assessment results are used to improve subsequent learning.

Assessment is a repeating cycle that involves the following basic actions:

- Making faculty expectations of student learning explicit and public
- Setting appropriate criteria and high standards for learning quality
- Systematically gathering, analyzing and interpreting evidence to determine how well performance matches those expectations and standards, and
- Using the resulting information to document, explain, and improve program performance.

The *purpose* of assessment activities is fourfold:

➢ To determine whether or not intended outcomes are being achieved and validate the need to undertake and continue certain initiatives
➢ To inform departmental faculty and other internal and external decision-makers about the relevant issues that can impact the program and student learning
➢ To provide information that can be used to focus conversations about policy, programs, and practices and determine how the policies, programs, and practices can be improved, and

➢ *Ongoing*. Assessment should be part of the ongoing business of the department, not only a priority during program review cycles or prior to accreditation visits.
➢ *Implemented gradually*. Assessment needs to become part of the culture slowly, implemented in carefully orchestrated steps over time.
➢ *Multi-faceted*. An effective assessment program utilizes multiple methods to assess each learning outcome, assesses multiple samples, and assesses students in different phases of learning.
➢ *Pragmatic*. To be truly useful, department level assessment must stay on a practical level with obvious implications to faculty and students.
➢ *Faculty-designed* and implemented.
➢ *Self-renewing.* Assessment data and information must feed back into the system, both on the university and department level.

What Assessment Is Not?

➢ It is not solely an administrative activity. Faculty must not merely tolerate or endorse assessment, but they must actively engage in it.
➢ Assessment should not be a part of an institution's faculty evaluation system. It should promote self-examination, critical questioning, evaluation, accountability, and renewal, but it should not punish individuals or programs honestly seeking to improve.
➢ Assessment is not intrusion into a faculty member's classroom, nor does it infringe on academic freedom.
➢ It is not necessarily testing, nor a series of tests. Testing can be part of assessment.
➢ Assessment is not quick or easy. It is conceptually, educationally, politically, and administratively complicated business.

ASSESSMENT AS SCHOLARSHIP

The motto of NSU's Office of Institutional Effectiveness and Assessment is "Assessment as Scholarship and Learning." We believe that effective assessment process follows general steps of good applied or action research. Five elements of research are essential for effective assessment:

1. Asking good questions

2. Identifying appropriate methods

3. Identifying/Developing reliable measures /instruments

4. Measuring the phenomena / Administering instruments

5. Communicating and using results effectively.

This manual is structured around these basic elements of good research. The flowchart on the next two pages outlines further discussion.

```
                    ┌─────────────────────────┐
                    │  Asking Good Questions   │◄────────────┐
                    └─────────────────────────┘              │
                               │                             │
                               ▼                             │
                      ┌─────────────────┐                    │
                      │ Mission of the  │                    │
                      │    program      │                    │
                      └─────────────────┘                    │
                               │                             │
                               ▼                             │
                      ┌─────────────────┐                    │
                      │  Program Goals  │                    │
                      └─────────────────┘                    │
                               │                             │
                               ▼                             │
                      ┌─────────────────┐                    │
                      │ Specific learning│                   │
                      │    outcomes     │                    │
                      └─────────────────┘                    │
     ┌──────────────┐         │                              │
     │ Validating   ├─────────┤                              │
     │ outcomes     │         ▼                              │
     └──────────────┘  ┌─────────────────────────┐           │
                       │ Identifying Appropriate │           │
                       │        Methods          │           │
                       └─────────────────────────┘           │
                               │                             │
     ┌──────────────┐          │                             │
     │ Linking      ├──────────┤                             │
     │ outcomes to  │          ▼                             │
     │ methods      │  ┌─────────────────────────────┐       │
     └──────────────┘  │ Identifying Reliable Measures/│      │
                       │       Instruments            │      │
                       └─────────────────────────────┘       │
     ┌──────────────┐          │                             │
     │ Surveys      ├──────────┤                             │
     └──────────────┘          │                             │
     ┌──────────────┐          │                             │
     │ Tests        ├──────────┤                             │
     └──────────────┘          │                             │
     ┌──────────────┐          │                             │
     │ Performance  ├──────────┤                             │
     └──────────────┘          ▼                             │
                      ┌─────────────────┐                    │
                      │   Standard      │                    │
                      │   Setting       │                    │
                      └─────────────────┘                    │
                               │                             │
                               ▼                             │
                       ┌─────────────────────────────┐       │
                       │ Measuring phenomena / Admin. │      │
                       │          design              │      │
                       └─────────────────────────────┘       │
     ┌──────────────┐          │                             │
     │ Sampling     ├──────────┤                             │
     └──────────────┘          │                             │
     ┌──────────────┐          │                             │
     │ Assessment   ├──────────┤                             │
     │ Plan         │          │                             │
     └──────────────┘          ▼                             │
                               ─────────────────────────────┘
```

6

```
                              │
                              ▼
                   ╭──────────────────────╮
                   │ Communicating Results │
                   ╰──────────────────────╯
                              │
   ┌─────────────┐            │
   │ Analysis of │────────────┤
   │ assessment data │         │
   └─────────────┘            │
   ┌─────────────┐            │
   │ Use of      │────────────┤
   │ assessment data │         │
   └─────────────┘            │
   ┌─────────────┐            │
   │ Reporting   │────────────┤
   └─────────────┘            │
                              ▼
                   ╭──────────────────────────────╮
                   │ Improvement / Accountability / │
                   │        Scholarship             │
                   ╰──────────────────────────────╯
                              │
                              ▼──────────────────────────▶
```

Communicating Results

Analysis of assessment data

Use of assessment data

Reporting

Improvement / Accountability / Scholarship

ASKING GOOD QUESTIONS

The first step in the assessment process involves asking the following questions.
- ➢ What is the purpose of our department / program?
- ➢ How does our department / program fit in NSU's institutional context?
- ➢ What do we want our students learn?
- ➢ Does our curriculum and teaching methods reflect our goals?

Overall Purpose or Mission of the Program / Department

The following questions may be useful for getting started with writing a mission statement:
- Whom do we serve?
- How do we serve?
- What results do we want to achieve?
- What is our guiding philosophy?
- How does our department/program fit into institutional, disciplinary, professional, regional context?
- What makes our department/program unique in its context?
- What future plans do we envision?

Elements of a good mission statement:

- ➢ *Focus problem* – What are the focus problems that we exist to solve? Who is affected by the problem? How would the community be improved, be changed or be different if our department were successful in solving the problem or responded to the need?
- ➢ *Purpose* – a short, succinct statement that describes the *end result* your department seeks to accomplish and to whom.
- ➢ *Primary means* – one sentence that describes the primary means used to accomplish the purpose. How do we accomplish our purpose?
- ➢ *Values* – a list of fundamental values and beliefs or guiding principles shared by members of the department and practices in their day-to-day work and interaction with others (We believe…; We assume…).

The goal of this process is not simply to write a mission statement but also achieve consensus about the purpose of the program. If the mission statement is written well and validated by consensus, it can become a focal document in the program.

Program Goals

Program goal is an intended outcome of instruction that has been stated in general enough terms to encompass a *domain* of student performance (e.g., "Graduates of the program will demonstrate good problem-solving skills"). A program goal must be further defined by a set of *specific* (observable and measurable) learning outcomes to clarify instructional intent.

Functions **of program goals**:

- They make clear the variety of types of learning expected from the instruction (i.e., knowledge, comprehension, performance skills, et.)
- They provide a focus for instruction that avoids concentrating on isolated and unrelated learning tasks
- They are general enough to permit flexibility in choosing teaching methods and materials
- They provide a framework for planning and preparing assessments and for interpreting assessment results.

Program goals generally are built upon the three basic categories of learning outcomes:

- ➢ *Cognitive outcomes* or what students know.
  - *Knowledge* outcomes are the lowest level of cognitive outcomes and are concerned with the recall or recognition of learned material
  - *Comprehension* is concerned with grasping the meaning of material as shown by interpretation, translation, prediction, and similar responses
  - *Application* is the ability to consciously use the material in new situations

- ➢ *Affective / attitudinal / value-based outcomes* or what students care about
  - Affective outcomes are concerned with feelings and emotions that are described by individual's disposition, willingness, preferences, enjoyments, and similar terms
    - attitudes
    - interests
    - appreciations
    - adjustments

- ➢ *Skill-based / behavioral / performance outcomes* or what students can do. For example:
  - *Skilled performance*: speaking, reading, singing, etc.
  - *Higher level skills*: lab skills, communications skills, specialized performance skills (e.g., in business, education, art)
  - *Critical thinking skills* emphasize analysis and evaluation (e.g., identifying and analyzing a problem; evaluating possible solutions, etc.)
  - *Creative thinking skills* emphasize production of something new (e.g., producing a plan for solving a problem)

Good Practices in developing program goals:

➢ *Have open discussions with department faculty on one of the following topics*:
  - Describe the ideal student in your program at various phases throughout your program. Ask:
    - What does this student know?
    - What can this student do?
    - What does this student care about?
    - List and briefly describe the program experiences that contribute most to the development of the ideal student
  - List the achievements you implicitly expect of graduates
  - Describe your alumni in terms of such achievements as career accomplishments, lifestyles, citizenship activities, etc.

➢ *Collect and review instructional materials*. Try sorting materials by the type of learning each one is designed to promote: recognition/recall, comprehension, application, higher order thinking skills, affective learning, skilled performance, etc.
  - Syllabi and course outlines
  - Course assignments and tests
  - Textbooks (esp. the tables of content and summaries)

➢ *Collect and review documents that describe your department and its programs*
  - Brochure and catalogues descriptions
  - Accreditation reports
  - Curriculum committee reports

➢ *Review and react to goals and outcomes from another unit that is similar but external*. Try grouping the statements into broad categories of student outcomes (e.g., knowledge, attitudes, behavior)

➢ Use the 25 percent problem to *refine or reduce a set of program goals*. Imagine that you want to reduce program or course material by 25 percent. What goals would you keep and which would you discard.

➢ *Use a Delphi technique*. Choose an impartial facilitator to mediate a panel discussion about possible program goals. In brainstorming session, ask each panel member to build a list of criteria that she/he thinks is important for program goals. For each criterion, have each member anonymously rank it as: 1-very important; 2-somewhat important; or 3-not important. Place the criteria in rank order and show the anonymous results to the panel. Discuss possible reasons for items with high standard deviations.

➢ *Use a theoretical frame of reference as a guide*. It is helpful to refer to some frame of reference that clarifies the various types of learning outcomes. This might provide greater assurance that important goals are not overlooked and provide a guide for stating and organizing the goals.
  - *Taxonomy of Educational Goals*

- Cognitive goals: concerned with intellectual outcomes; the classification system ranges from lower-level knowledge outcomes to higher-level abilities and skills (See Attachment 1)
- Affective goals: concerned with outcomes in the areas of interests, attitudes, appreciation, and methods of adjustment. Classification system ranges from receiving stimuli to developing a characteristic set of values that direct behavior.

- *Types of learning areas*. The list (see Attachment 2) delineates many of the major learning areas in which program goals might be produced
- *Types of student learning*. There are four general types of student learning that shape program goals both structurally and functionally:
    - Understanding (demonstration of what student *knows* and understands within a *specific* context)
    - Performance (demonstration of what student can *do* within a *specific* context)
    - Capability (demonstration of what student can *do* that can be *transferred* from one context to another)
    - Perspective (a student's demonstration of what student *knows* and understands that can be *transferred* from one context to another.)

➢ Worksheet below might be helpful for identifying educational goals of your department / program.

Program Goals Definition Worksheet

Each faculty member in the department should complete a copy of this worksheet. Arrange a time for all of you to sit down together to compare notes and discuss results. The final product of this exercise should be a list of three to five broad goals that describe what department faculty believe should be characteristic of graduates in the major.

1. List any department goals that you know. This information can most likely be found in the course catalog, program brochure, or department mission statement.

2. Describe your ideal student in terms of strengths, skills, knowledge and values, and identify which of these characteristics are the result of the program experience.

3. Keeping this ideal student in mind, ask what the student
     a. knows

---

b. can do

c. cares about

4. What program experiences can you identify as making the most contribution to producing and supporting the ideal student?

5. What should every graduate of your program know?

6. What career achievements of your alumni are you most proud of?

Considerations in selecting program goals:

➢ Do the program goals reflect orthogonal or institution-wide goals and the program's mission?
➢ Do the goals represent all (cognitive, affective, and behavioral) logical learning outcomes of your instructional area?
➢ Are the goals realistic and attainable by the students?
➢ Are the goals in harmony with basic principles of learning? Some of the basic factors that should be considered are the following:
  • *Readiness*: Do the students have the necessary experiences and educational background to proceed successfully?
  • *Motivation*: Do these goals reflect the needs and interest of the students?
  • *Retention*: Do these goals reflect learning outcomes that tend to be retained longest (e.g., comprehension, application, thinking skills)?
  • *Transfer value*: Do these goals reflect learning outcomes that are widely applicable to new situations? Do the goals reflect realistic and complex learning tasks that are most useful in the "real world"?

Specific Learning Outcomes

Views about academic quality and effectiveness have shifted over the past three decades from an almost exclusive pre-occupation with inputs (student and faculty credentials and resources) and processes (offerings, requirements, teaching loads, class size, student rating of instruction, time to degree, etc.) to a more mission specific focus on teaching and learning *outcomes*.

What exactly will our students know? What exactly will they understand? What exactly will they be able to do with their knowledge at the end of the program? Program faculty members should answer these questions in the framework of the program goals with sentences describing competencies that program graduates should possess or, in other words, the kinds of things that students know or are able to do after instruction that they did not know or could not do before.

 Specific Learning Outcome is an *intended* outcome of college experience that has been stated in terms of *specific* and *observable* and *measurable* student performance (e.g., "Students will be able to identify details that are explicitly stated in a passage"). Specific learning outcomes transform the general program goals into specific student performances and behaviors that demonstrate student learning and skill development along these goals.

In other words, specific learning outcomes describe the types of performance that learners will be able to exhibit when the program has achieved its goals. Specific learning outcomes provide an operational definition of what we mean when we identify program goals. Unless the general goals are further operationalized in this way, they will not provide an adequate framework for teaching, learning, and assessment.

For example,
1. Program graduates will be able to demonstrate solid problem-solving skills (*Program goal)*
   1.1. Students will be able to analyze a situation to identify a problem *(Outcome 1)*
   1.2. Students will use multiple resources to gain additional information regarding the problem *(Outcome 2)*
   1.3. Students will develop a procedure to solve the problem using a sufficient knowledge base *(Outcome 3)*
   1.4. Students will propose and critique a viable solution to the problem *(Outcome 4)*

Benefits of formulating intended learning outcomes are fourfold. Outcomes statements

- Form the operational basis of assessment at the course, program, and institutional levels
- Provide direction for all institutional activity
- Inform students about the intentions of the faculty
- Inform external stakeholders about the educational experience in a given program or department.

Components of Statements Specific Learning Outcomes:

➢ Essential Components:
- *Behavior*: specify actions or behaviors that follow instruction and could serve as evidence that the goal has been achieved
  e.g., communicate results
- *Object*: identify the focus of learning – content, concept(s), skill, or attitude
  e.g., public opinion polls

➢ Optional Components:
- *Target groups*: specify subgroups when goal applies differentially
  e.g., graduating journalism majors
- *Conditions*: give information about situations in which the student will be required to demonstrate the behavior – how, when, or where
  e.g., after analyzing and interpreting information
- *Performance criteria*: state a minimum level of performance
  e.g., in written, oral, and graphic forms
- *Performance stability*: give information about how often the student behavior must be observed to be a true indicator that the behavior is a stable part of the student's achievement repertoire
  e.g., at least three different groups.

Example:
"After analyzing and interpreting information from public opinion polls, the graduating journalism major will be able to communicate the results to at least three different groups in written, oral, and graphic forms."

Good statements of intended learning outcomes:

- *Are student-focused rather than instructor-focused*. Intended outcomes are formulated to focus on student learning, i.e. they describe what students should know, understand, or be able to do with their knowledge at the end of a program.
  Poor: "The program will include instruction in multimedia techniques"
  Good: "Graduates of the program will be able to use multimedia to prepare presentations

- *Focus on the learning resulting from an activity rather than on the activity itself*.
  Poor: "Students will study at least one non-literary genre of art"
  Good: "Students will arrive at an analytical and reasoned appreciation of a specific art form", "Students will be able to communicate the appreciation to others either in written or verbal form"

- *Reflect state mandates and institutional expectations about learning*. Typically these expectations address the transferable or orthogonal competencies (e.g., writing, critical thinking, leadership skills, quantitative reasoning.) Departments and programs should reinforce these broad goals in the statements of expected learning outcomes and, subsequently, in the curricula.

- *Are reflected in program curriculum and translated into course specific objectives*. A good practice is to ask instructors to state explicitly in each course syllabus the program level goals and outcomes addressed in that course.

- *Focus on important, non-trivial aspects of learning that are credible to the public.* One pitfall to avoid in formulating intended outcomes is focusing on easy-to-measure, but relatively unimportant outcomes like, "Students will recall the stages of mitosis." This can happen when learning outcomes are developed by carving up the content of the discipline into smaller pieces. The focus of learning outcomes is not on less content but rather is on what students can do with the content they have learned.
  For example, "Students will be able to reason effectively by using simplified economic models such as supply and demand, marginal analysis, benefit-cost analysis, and comparative advantage"

- *Are general enough to capture important learning but clear and specific enough to be measurable.* For example, the outcome, "Students will be able to solve problems," gives little guidance for assessment. In contrast, the outcome "Students will work effectively with others on complex, issue-laden problems requiring holistic problem solving approaches," can be assessed by developing assessments that require teams of students to develop solutions to complex, issue-laden problems, as defined by the discipline. They can, then, be judged on the effectiveness of their team skills, the quality of their solution, and their ability to use holistic problem solving approaches.

- *Are effectively worded*
    - use action verbs that describe definite, observable actions. Faculty members should select those verbs that (i) most clearly convey instructional intent and (ii) most precisely specify the student performance the program is willing to accept as evidence that the general instructional goal has been achieved (See Attachment 1, Section "Related Behaviors")
    - include a description under which the action takes place – "when given x, the student will be able to…"
    - indicate an appropriate level of competency that is assessable through one or more indicators.

Specific Learning Outcomes Worksheet

This worksheet may help you and others in your department develop specific learning outcomes from the program goals you have identified. Have all faculty members complete the following table. Meet as a group to discuss your response and try to reach consensus on desired goals and outcomes.

Program general educational goals	Specific learning outcome(s)
1.	a) b) c)
2.	a) b) c)
3.	a) b) c)
4.	a) b) c)

Validating Specific Learning Outcomes: Curriculum mapping

If students are expected to attain specified intended outcomes, they should be *provided with the opportunities to learn what they need to learn*. The purpose of curriculum mapping is to look at the program/department curriculum in light of intended outcomes to ensure that students receive appropriate instruction and assignments in the desired order and with enough repetition so that learning outcomes are achieved. Curriculum mapping enables a department or program to identify gaps in the curriculum as well as to provide an overview of what each course is attempting to accomplish. Curriculum mapping can be done in two general formats – curriculum audit grid and course objectives/program goal reference matrix.

Curriculum Audit Grid. Program Outcomes / Program Curriculum

Course					
Outcomes	201	203	207	320	425
1. Recognize and articulate approaches to psychology	I	E		R	R
2. Independently design valid correlational research		I		U	R
3. etc.					

Introduced=I, Emphasized=E, Utilized=U, Reinforced=R

Course Objectives / Program Goals Reference Matrix

Course:_____   Prepared by: _____Date:_____

Course Learning Objectives	Program Goals (Enter level number defined below)							
1	A=	B=	C=	D=	E=	F=	G=	H=
What will be measured to demonstrate that this learning objective has been achieved?								
2	A=	B=	C=	D=	E=	F=	G=	H=
What will be measured to demonstrate that this learning objective has been achieved?								
3	A=	B=	C=	D=	E=	F=	G=	H=
What will be measured to demonstrate that this learning objective has been achieved?								

A. – H. – departmental or program goals

Level:
1. Major emphasis of the course
2. Discussed in the course and covered in homework or quiz
3. Mentioned in the course but not covered in homework or quiz
4. Not mentioned in the course

IDENTIFYING APPROPRIATE METHODS

Basic rules:

- There should be at least one method for assessing each educational outcome, and no assessment that does not address at least one outcome should be included
- Multiple methods to assess the same outcome should be used whenever feasible.
- The ability of assessment to improve student learning depends on the relevance and usefulness of the information that is generated. To be useful, assessment methods must gather evidence that is closely related to the selected intended learning outcomes. When choosing assessment methods, make sure the selected methods
- answer questions that are important to you
- follow identified "good practices" in undergraduate education
- are manageable given available resources and skills
- result in useful feedback that highlights accomplishments and identifies areas requiring attention.

Taxonomy of Assessment Methods

Assessment methods can be categorized according to three taxonomies:
- selected-response/constructed-response
- direct/indirect, and
- qualitative/quantitative methods.

Selected-response assessments, also called the recognition assessment, present alternative responses, from which the student chooses the correct or preferred answer. Typical selected-response types are multiple-choice, true-false, and matching tests.

Constructed-response assessments, also called the production assessment, require students to produce an answer or furnish an "authentic" response to a given stimulus or test question. Typical constructed-response formats are sentence-completion tests, essay questions, and performances.

Direct methods require students to display their knowledge and skills as they respond to the instrument itself. Some examples of direct methods are:
- objective tests
- essays
- presentations
- classroom assignments

Indirect methods ask students to reflect on their learning rather than demonstrate it. Some examples of indirect methods are:
- surveys
- interviews
- focus groups

Quantitative studies assess teaching and learning by collecting and analyzing numeric data using statistical techniques. Researchers typically work with a small number of predetermined response categories. Emphasis is usually on analyzing a large number of cases using instruments that have been evaluated for their validity and reliability. Some examples of quantitative measures and techniques are:
- GPA
- grades
- exam scores
- tests
- structured interviews
- forced-choice surveys

Qualitative studies rely on descriptions rather than numbers and usually deal with unknown causes, variables, and an absence of explanatory theories. Qualitative approaches rely on discovery, subjectivity, and interpretation. Qualitative studies are useful
- for discovery when we do not know enough to formulate a hypothesis, and
- for communicating results of quantitative studies. Qualitative studies are helpful ways to illustrate and explain outcome results when numbers simply do not portray the meaning of the collegiate experience

However, qualitative assessment has some difficulties: it is hard to identify standards, qualitative assessment depends upon objectivity, which may be hard to find, and it suffers from lack of consistency or reliability in judgment between evaluators and over time.

Some examples of qualitative measures are:
- exit interviews
- writing samples
- open-ended questions on surveys.

Overview of Assessment Methods

Assessment Strategies

Data	Assessment Tool	Who or What is Analyzed?	What can be Assessed?
Self-reports	Classroom assessment Focus groups Interviews Reflective essays Surveys	Alumni Employers Enrolled students Faculty Graduating students Entering students Staff	*Perceptions about*: Campus climate Perceived learning Value-added Attitudes Values
Achievement tests	Test score analysis Content analysis Scoring rubrics	Embedded questions on exams Standardized tests Locally developed exams Oral thesis defenses	Mastery and knowledge of principles Skills Value-added
Observations	Case studies Observations	Campus events Classes Faculty offices Student services offices	Attitudes Interactions Student involvement Student learning
Student academic work	Content analysis Scoring rubric	Capstone course products Homework papers Term papers/projects Portfolios Presentations	Mastery and knowledge of principles, skills Values Processes Value-added
Campus documents	Course x program goals/outcomes matrix Course assignment x program goals/outcomes matrix Content analysis Analysis of forms	Administrative units Departments Programs Course syllabi Student transcripts	Accuracy Cohesion/consistency Efficiency Objectives Processes

The following section describes the basics for doing outcomes assessment using a variety of techniques. For each technique, procedures, strengths, and weaknesses are described. References for each technique are provided in the *Bibliography* section. In Attachment 5, additional cautions/advice for and some illustrative examples of selected methods could be found.

Advantages and Disadvantages of Assessment Methods

Method	Description	Strengths	Weaknesses
Alumni Survey	Surveying department alumni can provide a wide variety of information about program satisfaction, how well students are prepared for their careers, what types of jobs or graduate degrees majors have gone on to obtain, starting salaries for graduates, and the skills that are needed to succeed in the job market or in graduate study. These surveys provide the opportunity to collect data on which areas of the program should be changed, altered, improved or expanded.	Alumni surveying is usually a relatively inexpensive way to collect program data from individuals who have a vested interest in helping you improve your program as well as offering the opportunity for improving and continuing department relationships with program graduates.	Without an easily accessible and up-to-date directory of alumni, they can be difficult to locate. It also takes time to develop an effective survey and ensure an acceptable response rate.
Culminating Assignments	Culminating assignments offer students the opportunity to put together the knowledge and skills they have acquired in the major, provide a final common experience for majors, and offer faculty a way to assess student achievement across a number of discipline-specific areas. Culminating assignments are generally designed for seniors in a major or field to complete in the last semester before graduation. Their purpose is to integrate knowledge, concepts and skills that students are expected to have acquired in the program during the course of their study. This is obviously a curricular structure as well as an assessment technique and may consist of a single culminating course (a "capstone" course) or a small group of courses designed to measure competencies of students who are completing the program. A senior assignment is a final culminating project for graduating seniors such as a performance portfolio or a thesis that has the same integrative purpose as the capstone course.	Many colleges and universities are using capstone courses to collect data on student learning in a specific major or in general education or core requirement programs.	Putting together an effective and comprehensive capstone course can be a challenge, however, particularly for those programs that mesh hands-on technical skills with less easily measurable learning outcomes. Also, there is a great deal of start-up time to developing appropriate and systematic methods for assessing these or other culminating experiences.
Content Analysis	Content analysis is a technique that looks at a group of students, such as majors in a program or department, and assesses samples of written work that are produced by this group. This assessment method uses outcomes identified as important prior to the analysis or as the analysis proceeds. For example, you might want to determine how well majors in your department write. To use content analysis to assess their	Content analysis allows you to assess learning outcomes over a period of time and can be based on products that were not created for program assessment purposes. Because writing samples can be re-examined, content analysis also makes it easier to repeat portions of the study	Accuracy of the assessment is limited to the skill of the person(s) doing the analysis. Data is also limited by the set of written work and may not be relevant to technical skills valued by a particular field or major that involve hands-on performance. Pre-testing coding schemes,

Advantages and Disadvantages of Assessment Methods

Method	Description	Strengths	Weaknesses
	writing skills, you will need a representative sample of the writing. Analysis may look at what students actually write or at the underlying meaning of their writing. Results are generally presented in written form giving averages and examples of specific categories of outcomes (e.g., spelling errors). Primary trait analysis, which identifies important characteristics of specific assignments and assigns levels of competency to each trait, can be particularly effective in identifying student learning.	and provides an unobtrusive way to assess student learning.	using more than one analyst per document, and concrete materials and coding schemes can improve the reliability of this technique.
Course - Embedded Assessment	Course-embedded assessment refers to methods of assessing student learning within the classroom environment, using course goals, objectives and content to gauge the extent of the learning that is taking place. This technique generates information about what and how students are learning within the program and classroom environment, using existing information that instructors routinely collect (test performance, short answer performance, quizzes, essays, etc.) or through assessment instruments introduced into a course specifically for the purpose of measuring student learning.	This method of assessment is often effective and easy to use because it builds on the curricular structure of the course and often does not require additional time for data collection since the data comes from existing assignments and course requirements.	Course-embedded assessment does, however, take some preparation and analysis time and, while well documented for improving individual courses, there is less documentation on its value for program assessment.
Curriculum Analysis	Curriculum analysis involves a systematic review of course syllabi, textbooks, exams, and other materials to help you clarify learning objectives, explore differences and similarities between course sections, and/or assess the effectiveness of instructional materials. It offers a way to document which courses will cover which objectives and helps in sequencing courses within a program. Also see Matrices.	Using curriculum analysis as an assessment tool can be a valuable way of tracking what is being taught where. It can provide assurance that specific learning goals and objectives are being covered in the program and can pinpoint areas where additional coverage is needed.	This method, however, can be time-consuming, particularly in large departments with many courses and different instructors, and there may be little consistency between how learning objectives are addressed in one course and how they are taught in another.
Delphi Technique	The Delphi technique is used to achieve consensus among differing points of view. In its original form, a team of experts, who never actually meet, are asked to comment on a particular issue or problem. Each member's response is	The Delphi technique can be useful in bringing together diverse opinions in a discussion forum.	This technique fails, however, when the facilitator lacks objectivity or when the participants feel unsafe or insecure in voicing their real opinions. For instance, a

Advantages and Disadvantages of Assessment Methods

Method	Description	Strengths	Weaknesses
	reviewed and a consensus determined. Any member whose response falls outside of the consensus is asked to either defend or rethink the response. The anonymity provided by this technique offers more junior members of the team an equal chance to get their ideas out, as well as permits a challenge to the ideas of senior members that might never take place in an open forum. More recently, the Delphi technique has been modified so that teams of individuals are brought together to discuss an issue or problem face-to-face and reaching a consensus at the meeting. For instance, a team of faculty members might meet to review possible goals and objectives for their department in an effort to develop a set of goals and objectives on which they can agree.		faculty member discussing intended goals and objectives might not be comfortable in disagreeing with the department head. For this technique to succeed, care must be taken to appoint an impartial facilitator and to convince participants that differing opinions are welcome.
Employer Survey	Employer surveys help the department determine if their graduates have the necessary job skills and if there are other skills that employers particularly value that graduates are not acquiring in the program. This type of assessment method can provide information about the curriculum, programs and student outcomes that other methods cannot: on-the-job, field-specific information about the application and value of the skills that the program offers.	Employer surveys provide external data that cannot be replicated on campus and can help faculty and students identify the relevance of educational programs.	As in any survey, ambiguous, poorly worded questions will generate problematic data. Additionally, though data collected this way may provide valuable information on current opinion, responses may not provide enough detail to make decisions about specific changes in the curriculum or program. Also, it is sometimes difficult to determine who should be surveyed, and obtaining an acceptable response rate can be cost and time intensive.
Focus Groups	Focus groups are structured discussions among homogeneous groups of 6-10 individuals who respond to specific open-ended questions designed to collect data about the beliefs, attitudes and experiences of those in the group. This is a form of group interview where a facilitator raises the topics for discussion and collects data on the results. Emphasis is on insights and ideas.	Focus groups can provide a wide variety of data about participants' experiences, attitudes, views and suggestions, and results can be easily understood and used. These groups allow a small number of individuals to discuss a specific topic in detail, in a non-threatening environment.	Data collected in this way, however, is not useful for quantitative results, and qualitative data can be time-consuming and difficult to analyze because of the large amount of non-standardized information. Ultimately, the success of this method depends on a skilled, unbiased moderator and appropriate groups of participants.

Advantages and Disadvantages of Assessment Methods

Method	Description	Strengths	Weaknesses
Institutional Data	A variety of departmental and student data are routinely collected at the university level. These data can enhance and elaborate on data you collect in the department. Institutional data can tell you whether the program is growing, what the grade point average is for majors in the program, and what the retention rate is for your students.	Institutional data are generally easily accessible and readily available. You can access this data through the Office of Institutional Effectiveness &Assessment and Office of Institutional Research. Student and departmental data are collected on a systematic and cyclical schedule that can offer you both current and longitudinal information.	These data sets are generally large and may be difficult to sort through, particularly for those individuals who are not used to working through large databases. The data may be less useful to specific departments or programs because the information collected is very often general (age, gender, race, etc.) and may not directly relate to program goals and objectives.
Matrices	At its most basic, a matrix is a grid of rows and columns used to organize information. For assessment purposes, a matrix can be used to summarize the relationship between program objectives and course syllabus objectives, course assignments, or courses in a program or department. Matrices can be used for curriculum review, to select assessment criteria or for test planning. A matrix can also be used to compare program outcomes to employer expectations.	Using a matrix can give you a good overview of how course components and curriculum link to program objectives, can help you tailor assignments to program objectives, and can lead to useful discussions that in turn lead to meaningful changes in courses or curricula.	However, because a matrix can offer a clear picture of how program components are interconnected and can reveal where they are not, acknowledging and responding to discrepancies may involve extensive discussion, flexibility and willingness to change.
Observations	Observation as a method of assessment is an unobtrusive tool that can yield significant information about how and why students learn. You may choose to observe any relevant interactive event, such as classes, club meetings, or social gatherings. This tool is generally used when you are interested in how students study, are concerned about the effectiveness of study sessions or other supplementary activities, or when you are focusing on the relationship between out-of-class behavior and in-class performance. Data collected through observation can be correlated with test scores and/or course grades to help provide further insight into student learning.	Data collected through observation can yield important insight into student behavior that may be difficult to gauge through other assessment methods. This method is typically designed to describe findings within a particular context and often allows for interaction between the researcher and students that can add depth to the information collected. It is especially useful for studying subtleties of attitudes and behavior.	Observed data, however, is not precise and cannot be generalized to larger populations. Conclusions may be suggestive rather than definitive, and others may feel that this method provides less reliable data than other collection methods.

Advantages and Disadvantages of Assessment Methods

Method	Description	Strengths	Weaknesses
Performance Assessment	Performance assessment uses student activities to assess skills and knowledge. These activities include class assignments, auditions, recitals, projects, presentations and similar tasks. At its most effective, performance assessment is linked to the curriculum and uses real samples of student work. This type of assessment generally requires students to use critical thinking and problem-solving skills within a context relevant to their field or major. The performance is rated by faculty or qualified observers and assessment data collected. The student receives feedback on the performance and evaluation.	Performance assessment can yield valuable insight into student learning and provides students with comprehensive information on improving their skills. Communication between faculty and students is often strengthened, and the opportunity for students' self-assessment is increased.	This type of assessment is labor-intensive, is sometimes separate from the daily routine of faculty and student, and may be seen as an intrusion or an additional burden. Articulating the skills that will be examined and specifying the criteria for evaluation may be both time-consuming and difficult.
Portfolios	Portfolios are collections of student work over time that are used to demonstrate student growth and achievement in identified areas. Portfolios can offer information about student learning, assess learning in general education and the major, and evaluate targeted areas of instruction and learning. A portfolio may contain all or some of the following: research papers, process reports, tests and exams, case studies, audiotapes, videotapes, personal essays, journals, self-evaluations and computational exercises. Portfolios are often useful and sometimes required for certification, licensure, or external accreditation reviews.	Portfolios not only demonstrate learning over time, but can be valuable resources when students apply to graduate school or for jobs. Portfolios also encourage students to take greater responsibility for their work and open lines of discussion between faculty and students and among faculty involved in the evaluation process.	Portfolios are, however, costly and time-consuming and require extended effort on the part of both students and faculty. Also, because portfolios contain multiple samples of student work, they are difficult to assess and to store and may, in some contexts, require too much time and effort from students and faculty alike.
Pre-test / Post-test Evaluation	This method of assessment uses locally developed and administered tests and exams at the beginning and end of a course or program in order to monitor student progression and learning across pre-defined periods of time. Results can be used to identify areas of skill deficiency and to track improvement within the assigned time frame. Tests used for assessment purposes are designed to collect data that can be used along with other institutional data to describe student achievement.	Pre-test/post-test evaluations can be an effective way to collect information on students when they enter and leave a particular program or course, and provide assessment data over a period of time. They can sample student knowledge quickly and allow comparisons between different students groups, or the same group over time.	Pre-test/post-test evaluations require additional time to develop and administer and can pose problems for data collection and storage. Care should be taken to ensure that the tests measure what they are intended to measure over time (and that they fit with program learning objectives) and that there is consistency in test items, administration and application of scoring standards.

Advantages and Disadvantages of Assessment Methods

Method	Description	Strengths	Weaknesses
Reflective Essays	Reflective essays may be used as an assessment tool to gauge how well students are understanding class content and issues. They are generally short essays (5 to 10 minutes) on topics related to the course curriculum and may be given as in-class assignments or homework. Reflective essays may be voluntary or required, open-ended questions on surveys required in student portfolios or capstone composition courses.	Reflective essays as an assessment tool can offer data on student opinions and perspectives at a particular moment in a class. Essays will provide a wide array of different responses and might lead to increased discussion among faculty and students.	Poorly worded, ambiguous questions will yield little, and opinions and perceptions may vary in accuracy. Analysis of essay content also takes additional time and expertise.
Standardized and Local Test Instruments	Selecting a standardized instrument (developed outside the institution for application to a wide group of students using national/regional norms and standards) or a locally-developed assessment tool (created within the institution, program or department for internal use only) depends on specific needs and available resources. Knowing what you want to measure is key to successful selection of standardized instruments, as is administering the assessment to a representative sample in order to develop local norms and standards. Locally-developed instruments can be tailored to measure specific performance expectations for a course or group of students.	Locally-developed instruments are directly linked to local curriculum and can identify student performance on a set of locally important criteria. Standardized tests are immediately available for administration and, therefore, are less expensive to develop than creating local tests from scratch. Changes in performance can be tracked and compared to norm groups and subjectivity/misinterpretation is reduced.	Putting together a local tool is time-consuming as is development of a scoring key/ method. There is also no comparison group and performance cannot be compared to state or national norms. Standardized measures may not link to local curricula and purchasing the tests can be expensive. Test scores may also not contain enough locally-relevant information to be useful.
Student Surveys and Exit Interviews	Surveys and interviews ask students to respond to a series of questions or statements about their academic experience. Questions can be both open-ended (respondents create answers) and close-ended (respondents answer from a list of simple and unambiguous responses). Surveys and interviews can be written or oral (face-to-face) or phone. Types of surveys include in-class questionnaires, mail questionnaires, telephone questionnaires, and interviews. Interviews include structured, in-person interviews and focus group interviews.	Surveys can be relatively inexpensive and easy to administer, can reach participants over a wide area, and are best suited for short and non-sensitive topics. They can give you a sense of what is happening at a given moment in time and can be used to track opinions. Data is reasonably easy to collect and tabulate. An interview can follow-up on evasive answers and explore topics in-depth, collecting rich data, new insights, and focused details.	Ambiguous, poorly written items and insufficient responses may not generate enough detail for decision making. Information may be distorted by the respondent, who may feel a lack of privacy and anonymity. The success of the interview depends ultimately on the skills of the interviewer.
Syllabus	Syllabus analysis (as well as systematic review of textbooks,	Use syllabus analysis when you want to	However, this review is time consuming

Advantages and Disadvantages of Assessment Methods

Method	Description	Strengths	Weaknesses
Analysis	exams and other curricular material) involves looking at the current course syllabus (written or oral assignments, readings, class discussions/projects and course expectations) to determine if the course is meeting the goals and objectives that the instructor or department has set for it.	clarify learning objectives; explore differences and similarities between sections of a course; or assess the effectiveness of instructional materials. Syllabus analysis can provide invaluable information to enhance any assessment plan.	and, as there may be more than one reviewer, there may not be adequate consistency in collecting and analyzing the data.
Transcript Analysis	Transcript analysis involves using data from student databases to explore course taking or grade patterns of students. This tool can give you a picture of students at a certain point in their academic careers, show you what classes students took and in what order, and identify patterns in student grades. In sum, transcript analysis gives you a more complete picture of students' actual curricular experiences. Specific information can be drawn from transcripts to help answer research questions, and course pattern sequences can be examined to see if there is a coherence to the order of courses taken.	Transcript analysis is an unobtrusive method for data collection using an existing student database. This information can be linked to other variables such as sex or major, or used to measure outcomes.	It is important to keep in mind, however, that course patterns may be influenced by other variables in students' lives that don't show up on their transcripts. Also, solutions that arise from results of the analysis may not be practical or easily implemented. It is critical to have specific questions whose answers can lead to realistic change before conducting the analysis.

Method Selection

Matrices

The following two matrices can greatly assist faculty members is selecting appropriate methods of assessment. The first example shows a set of criteria down the first column of the matrix and, along the first row, different assessment methods.

Selection Criteria Matrix

	Methods				
Criteria	Tests	Performances	Portfolios	Surveys	Classroom assignments
Match to the purposes of assessment					
Match to curriculum					
Technical quality					
Preparation time					
Costs					
Value to students					
Programmatic information					

The second example shows a set of educational goals down the first column of the matrix and, along the first row, different assessment methods.

Learning Goals by Methods Matrix.

	Methods		
Goals	Term paper	Questionnaire	Speech
Write at scholarly level	X		
Adapt verbal messages to a specific audience			X
Value life-long learning		X	

Attachment 3 provides examples of assessment methods in a variety of academic programs.

Guidelines for selecting assessment method

➢ Choose tools that evaluate intended learning outcomes. The outcomes for the program must govern the search for method.
➢ Choose method that yield manageable information. Do not let methods you choose make you an assessment slave.
➢ Use or modify course assignments. Chances are extremely good that you are already requiring students to do assignments that measure outcomes your program cares about.
➢ Use information that other campus units gather. Knowing when students enter into a major, what their academic aptitudes are, when they drop out of a major, where they

live after graduation, and where they are employed may all be available from other offices on campus.

➢ Choose methods that assess more than one outcome at a time.

➢ Network with other departments, the University-Wide Assessment Committee, and the Office of Institutional Effectiveness and Assessment so that your methods and tools and theirs are not redundant. If we can coordinate our efforts, we can save time and money.

➢ Use multiple methods that have different strengths and weaknesses. The multi-method approach yields a more valid view of students and also helps faculty evaluate the adequacy of their measures.

IDENTIFYING / DEVELOPING APPROPRIATE MEASURES / INSTRUMENTS

General Recommendations

Measurement means that outcomes are subject to some form of judgment (quantitative or qualitative) against criteria that can be applied with a high degree of reliability – i.e., repeatability and interjudge agreement.

Main criteria for evaluating the accuracy and appropriateness of measures of students' learning and development

> ➢ *Does the content of the measure address the question being asked?*

As a general rule, there should be a high degree of correspondence between what is being measured by an assessment instrument and the department / program's goals and intended outcomes of student learning (*content validity*).

Programs rushing into assessment activities frequently skip this step of determining content relevance and representativeness. Often, commercial tests are used for assessment without giving a thorough consideration how well that test fits the program's goals and curriculum. Before adopting a commercially available test, faculty members should match systematically each item from the test with the program's goals (e.g., by using a table of specifications.) If too many goals have too few items for representation, the test will have low content validity for the program.

> ➢ *Does the empirical structure of the data produced by the measure answer the question being asked?*

Questions about the structure of assessment data go to the heart of issues of reliability and validity. A variety of approaches can be used to evaluate the reliability of an assessment measure, and the choice of one approach over another should be based on what is being assessed and how.

If we are interested whether the items on the given assessment instrument measure the same characteristic, a measure of internal consistency is appropriate. If an answer is scored as correct or incorrect, internal consistency reliability is determined with the Kuder-Richardson-20 formula. If the instrument has more than one correct answer per test item, then Cronbach's coefficient of internal consistency would provide necessary reliability information.

When the assessment design is longitudinal and involves pre- and post-test, questions about consistency over time and estimates of test-retest reliability are appropriate. If assessment results are to be used for group or program purposes, reliability coefficients of .6 or above are generally acceptable. If one wishes to report results back to individual students, greater precision (>.7) is necessary.

Qualitative methods require a different approach to the evaluation of reliability. When, for example, multiple raters are used to assess students' writing samples, evaluations of interrater agreement are needed to ensure consistency in the evaluation process. When a scoring rubric is used, these measures should be supplemented with measures of the agreement between raters and "true" scores on anchor papers that represent categories or points along a continuum in the scoring rubric. Interrater reliabilities below .7 are suspect to their accuracy.

Direct evaluations of the structures of assessment data frequently focus on what is traditionally called *construct validity*. The question in hand here is whether the test or other design measures what it purports to measure or, in other words, whether there is congruence between the stated structure of a test and the actual structure of data from that test. Construct validity can, for example, be assessed by utilizing factor analysis, which examines correlations among scores from a test or survey in order to identify the factors underlying the instrument.

> ➢ *Are the data produced by the measure sensitive to students' educational experiences and the effects of educational intervention?*

Evaluating the sensitivity of scores to educational effects involves two related sets of analyses. The first set examines the relationships between scores and students' educational experiences, whereas the second examines the sensitivity of scores to factors, such as gender, entering ability, on-/off-campus status, that are not related to students' educational experiences.

Traditionally, correlations among assessment scores, educational experiences, and non-educational factors have been used to determine whether assessment scores are more strongly related to educational or non-educational factors. However, more complex procedures, such as differential item functioning, standardization, and item response models, become increasingly popular.

Guidelines to keep in mind when preparing assessment instruments from program goals / intended learning outcomes:

> ➢ Check to be sure that the specific learning outcomes include the most important elements of each goal / outcome
> ➢ Use the intended learning outcomes as criteria for preparing the assessment instrument
> ➢ Select the type of assessment instrument that best evaluates the knowledge, performance skills, and/or affective outcomes
> ➢ Prepare an instrument that is understandable and easy to use by students
> ➢ Use an appropriate scoring procedure (e.g., holistic for overall judgment, analytic for judgment on each criterion)
> ➢ Share the assessment instruments with students before instruction begins and have students use them to assess their own learning progress during instruction
> ➢ Have students participate in preparing the assessment instruments.

Development of Surveys

Types of Questions
Table below outlines main types of survey questions, their use, advantages and disadvantages.

Types of Questions	Use	Advantages	Disadvantages	Examples
Open Ended Questions	Stimulates free thought from the respondents. Elicits information that respondents can recall without difficulty when there are a very large number of possible answers and listing all of them as response choices makes answering the question difficult.	These questions stimulate free thought, solicit suggestions, probe people's memories, and clarify positions.	These questions require people to find the terms with which to express themselves. Answers may be incomplete, uninterruptible, or irrelevant. Information may be difficult to analyze.	1. What should be done in order to improve the department? 2. To what professional organizations do you belong?
Closed-Ended Questions With Ordered Answer Choices	Determines intensity of feeling, degree of involvement, and frequency of participation. Ordered choices provide specific limits to responses.	These questions are less demanding to answer. Answers may be combined to form a multiple-item scale.	The responses may not be exhaustive.	1. How many hours a week do you study? a. 0-3 hours b. 4-7 hours c. 8-11 hours d. 12-15 hours e. 16 or more hours
Closed-Ended Questions With Unordered Answer Choices	Provides independent choices representing different concepts.	Questions of this type are often used to establish priorities among issues and decide among alternative policies.	Preferred options of all respondents may not be stated. Respondents must balance several ideas at once, especially if asked to rank 10-20 items.	1. Rank in order of importance the following reasons for attending this university. - Reputation of the university - Reputation of the department - Close to home - Friends attend - The size of the university

Types of Questions	Use	Advantages	Disadvantages	Examples
Partially Close-Ended Questions	Provides for responses that might be overlooked by researchers.	These questions allow respondents to give answers when the available choices do fit them.	A sufficient number of additional responses to warrant analyses may not be obtained.	1. What are your plans for next year? a. Continue prior job b. Start a new job c. Continue graduate study d. Other

Common Response Items

Table below outlines main types of responses to survey questions, their use, advantages and disadvantages.

Types of Responses	Description	Advantages	Disadvantages	Examples
Likert Scale	This scale is used with attitude and opinion questions. Respondents are asked to indicate the degree to which they agree or disagree with statements. Statements are usually worded fairly strongly, and can be worded both positively and negatively.	Questions are easily understood and quantified. Undecided responses can be accommodated. Allows for depth of response. Provides a meaningful way to group a series of items. Overall scores can be computed.	Method is less direct than using some other answer categories that more closely match the questions.	1. General Studies classes are very important. __Strongly Agree __Agree __Undecided __Disagree __Strongly Disagree
Semantic Differential Scale	This scale is best used to describe a series of attitudes toward a complex concept. The question presents the topic or issue, and the semantic differential scale asks the respondent to choose a number between two opposite adjectives.	Generally strong at finding particularly favorable or objectionable aspects of multi-faceted issues and concepts. Provides an overall scale score (average) for the concept.	Limited applicability	1. Do you feel that computer instruction is: Efficient -3 -2 -1 0 1 2 3 Inefficient Useful -3 -2 -1 0 1 2 3 Useless Boring -3 -2 -1 0 1 2 3 Interesting

Types of Responses	Description	Advantages	Disadvantages	Examples
Importance	These categories can be used when you need to find the importance of goals, objectives, or activities.	These categories simplify analysis, survey construction, and data entry. Works well to force respondents to give meaningful replies.	Not much variance is allowed, and statistical uses are limited. Does not provide much discrimination between items.	1. How important is…? __Very Important __Somewhat Important __Not Important
Participation	This scale can be used for questions about the frequency of activities, when respondents aren't likely to know the actual number of times they participated.	Easy to complete, and works well to describe several activities which would have a wide range of expected frequencies.	Not very precise. Use numeric ranges such as 1-5, 6-10, etc. for additional precision, if the items have similar expected ranges.	1. How often did you attend…? __Very Often __Often __Sometimes __Rarely __Never

When constructing a survey, make sure survey items and response categories:

- Do not lead the respondent into giving an answer that he/she would not ordinarily give. The survey items should be stated in a neutral manner
- May not be interpreted in more than one way
- Are not vague
- Are not too personal
- Are not likely to be endorsed by almost everyone or almost no one
- Have words that are simple, clear, direct, and uniformly understood
- Are short, rarely exceeding 20 words
- Contain only one question – avoid double-barreled questions
- Avoid, if possible, universals such as "all", "always", and "none"
- Avoid, if possible, words such as "only", "just", "merely", and others of a similar nature
- Are in the form of simple sentences rather than compound or complex sentences
- Avoid the use of double negatives
- Are not hypothetical
- Avoid abbreviations or unconventional phrases
- Avoid unequal comparisons among response categories.

Achievement Testing: Development of Instruments

An achievement test is a device for measuring a sample of student performance. To provide a valid measure, the test must be relevant and representative of what has been taught during the instruction. The following steps provide a useful procedure for test preparation:

1. *Prepare a table of specifications*
2. *Construct relevant test items*
3. *Prepare a file of test items*
4. *Prepare detailed specifications for items*

Table of Specifications

A table of specifications is a two-dimensional table that relates the instructional goals and intended learning outcomes to the course content. A completed table describes the number of test items needed to obtain a balanced measure of the instructional goals and the course content emphasized in the instruction.

A sample table of specifications for a 50-item test in Economics.

	General learning Outcomes			
	1. Knows basic terms	*2. Comprehends concepts and principles*	*3. Applies principles*	*4. Interprets data*
Content Areas				
A. Forms and functions of money	3	4	3	
B. Operation of banks	4	3	5	3
C. Role of Federal Reserve System	4	6	3	2
D. State regulation of banks	4	2	4	
Total number of test items	15	15	15	5

The numbers listed with the table indicate the number of test items to be constructed in each area. The relative emphasis should reflect the emphasis given during development of intended learning outcomes and during instruction. This is accomplished by assigning weights to each outcome and to each content area during the construction of the table. The usual procedure is first to distribute the *total* number (or percentage) of test items over the outcomes and content areas and then to distribute the items among the individual points.

Constructing relevant test items

For example, to construct one of the four test items to measure the first goal in content B in the table of specifications above, we would take the following steps:

1. Select one of the specific learning outcomes listed under the general instructional goals
2. Select one of the important banking terms
3. Construct a test item that calls forth the specific performance indicated in the learning outcome.
Example:
Instructional Goal: 1. Knows basic terms
Learning Outcome: 1.1. Relates terms that have the same meaning
   1. Checking accounts are also called
      *A. demand deposits
       B. time deposits
       C. currency
       D. credit money.

Suggestions for Item Writing (All Item Types)

- Write items directly after instruction when possible.
- Write more items than are needed – 25% more is a good rule of thumb.
- Keep reading and vocabulary levels simple unless you are testing those skills (i.e., avoid superfluous wording).
- Make sure the problems posed in the item stems[1] are clear and unambiguous.
- Use "concrete" situations and pictorial, graphic, or tabular stimuli when possible.
- Use novel material in formulating problems to measure higher-level mental processes (i.e., analysis and application type items); however, guard against overuse of this strategy.
- Avoid lifting statements verbatim from a text unless your intention is to test memorization skill or knowledge for simple and basic information.
- Avoid single and double negative items; phrase in a positive format if possible.
- Be careful of clues and answer cues within the item stem.

   Item Responses

- Avoid, if possible, noun modifiers like "all", "never", "no", and "always"
- Order the response options in some parallel way (e.g., alphabetically, chronologically), but avoid creating response patterns (e.g., an alternating true-false pattern, overuse of the "C" option in a set of multiple-choice items).

Suggestions/Checks for Multiple-Choice Items

- Avoid highly technical response options.

---

[1] Item stem is the part of a test item that poses the question or sets up the problem situation; the stimulus.

- Avoid having the correct answer longer than the incorrect options.
- Use responses that are plausible and homogeneous in some way.
- Use between three and five options. Try to use the same number of options for all items; however, do not create superfluous options just to maintain a parallel format.

Suggestions/Checks for Matching Items

- Provide more choices than number of statements to be answered unless a choice can be used more than once.
- Have students choose answers from the column with the least amount of reading.

Suggestions/Checks for True-False (Alternative Response) Items

- Avoid ambiguous and indefinite terms of degree or amount (i.e., "frequently", "in most cases", etc.).
- Avoid negative and double negative statements.
- Keep true and false statements approximately the same length.
- Have approximately the same number of true and false items on a test.

Suggestions/Checks for Writing Completion or Short Answer Items

- Draft items that require a single-word answer or a brief and definitive statement.
- Avoid statements that may be logically answered by several terms.
- Indicate the unit of expression (i.e., date, percentage) when answers require numerical information.

Suggestions/Checks for Writing Essay Items

- Select items carefully because of the limited number that can be given in a single time frame.
- Make items clear and specific so that scoring can be done easily.
- Establish a framework within which the student will write
    - Limit the area covered by an item.
    - Indicate the value of items and suggest time parameters.
    - Decide, in advance, the factors considered in evaluation and note them in the instructions

Preparing a file of test items and using detailed specifications

An item bank is a large collection of test items from which an instructor can draw items that are matched to a specific measurement need or purpose. When stored as an electronic database, the item bank enhances the practice of academic measurement. A file of test items (an item bank, an item pool) can easily be prepared by recording each item on a separate card and placing above it the content area, goal, and specific learning outcome being measured. When test items are being prepared cooperatively by instructors, detailed specifications may be useful for item writing. In this case, it is a good practice to

- state the general program goals
- state the specific learning outcome
- indicate the type of test item to be used (e.g., multiple choice)
- describe characteristics of the item (item stem, correct response, incorrect alternatives)

The use of detailed specifications provide greater assurances that a functionally equivalent set of relevant test items will be prepared for each specific learning outcome, despite the fact that several instructors are constructing the items.

Good items banks have a number of distinguishing features:
- The number and type of items faithfully reflect the nature and emphases of the knowledge domain to be measured. More so than paper-and-pencil testing, modern computer-based testing requires a sufficient number of interchangeable items on each test objective for multiple-form and adapted tests
- The items meet accepted standards of content validity and psychometric quality. Simply put, the items measure what they are supposed to measure and they do so very well
- The item bank is easy to use and maintain. Content specialists such as teaching faculty can easily manage the test item database and build tests to their specifications.

The development of a sound item database is an arduous and lengthy task. The rewards, however, in terms of more precise and efficient measurement are substantial.

Standard Setting for Achievement Testing

There are two basic ways to establish standards for program assessment:

➢ *Criterion-referenced assessment* – To what degree did we accomplish our intended educational outcomes? Criterion-reference assessment reports how well students are doing relative to predetermined performance level on a specified set of learning outcomes.

➢ *Norm-referenced assessment* – How does performance at our program compares to performance at other institutions or programs? The main goal of norm-referenced assessment is to classify students. This type of assessment is designed to highlight achievement differences between and among programs. Results of norm-referenced assessment are interpreted in relation to the performance of a large group of similar students who took the test when it was first normed. The information provided by this type of assessment can be useful for deciding how the program is compared to similar programs in peer institutions, for deciding whether or not students need remedial assistance, etc. However, this type of information gives little information about how the program performs or about what the student actually knows, values or can do.

The further discussion will be primarily focused on the criterion-referenced assessment because this type of assessment provides more useful information for assessing program effectiveness and for program improvement. To determine criterions or performance levels,

cut points must be established that reflect what students in each performance levels should know or able to do. In other words, cut points are set such that students whose performance exceeds the particular cut point are inferred to possess sufficient knowledge, abilities, and skills to be regarded as performing at that level relative to the academic standards. Two most popular standard-setting procedures for the specifications of performance levels on criterion-based assessment are (i) modified Angoff method and (ii) Bookmark procedure.

Steps in the Modified Angoff Method

1. Group of expert faculty members (panelists) draft descriptions of achievement levels
2. Panelists judge each test item in order of test administration and give their estimate of the difficulty level of each item
3. Each judge is then asked to look at one item at a time and give the best estimate of the proportion of a hypothetical group of borderline or "barely proficient" students that are expected to answer the item correctly
4. These individual judgments are then discussed among the panelists[2]
5. Each judge is then asked to give another estimate of the difficulty level of each item
6. Panel discusses these individual judgments again to produce concurrence of the judgments. This discussion is supposed to produce approval of the standards set by the panelists
7. The judgmental *p*-values (or item difficulties) for each item are then aggregated; disagreement among the experts is usually resolved by applying some type of averaging technique.

Advantages of the modified Angoff method:
- Angoff method does not require *actual* test data
- It is relatively easy to use
- It is widely accepted.

Disadvantages of the modified Angoff method:
- This method is inappropriate for use with constructed-response items. The Angoff method can be used for selected-response (multiple choice) items *only*
- Requires judgments of item difficulty levels that are very difficult to make.

Steps in the Bookmark Procedure

1. Using *actual* test data (*p*-values or item difficulties[3]), test items are ordered from least difficult to most difficult and compiled into a booklet. Typically, the items are ordered based on the proficiency required for an examinee to have 2/3 probability of answering the item correctly after guessing was factored out.
2. Panelists (expert faculty) are directed to start with lest difficult item and move through booklet until they find the place where hypothetical "barely proficient" student would

---

[2] Actual student performance on the test (item difficulty or *p*-values) may be introduced at this point
[3] Simply, the difficulty index is the number of students who answered the item correctly divided by the total number of students.

probably get all items to that point correct and probably get all items beyond that point incorrect. At that point in the booklet, the faculty members place their bookmarks

3. Panelists discuss the reasons for their individual selections of bookmarks
4. The actual performance data (proportion answering each item correctly and the cumulative percent of students at each score point) are provided to the faculty panelists
5. After seeing the data, panelists are asked to make second bookmark placement for how they expect the "barely proficient" students to perform. The second bookmark could be either the same or different from the initial bookmark
6. The cut score is based on the second bookmark placement and is calculated by summing, for each panelist, the number of items up to the bookmark and then averaging those values across the panelists
7. Descriptors for performance levels are then written.

Advantages of the Bookmark procedure:
- The Bookmark method can handle dichotomously scored selected-response items and polytomously scored constructed-response items
- This procedure simplifies the judgmental tasks.

Disadvantages of the Bookmark procedure:
- The Bookmark method requires actual student results on either the item pool or the test forms
- This procedure requires knowledge of principles of item-response theory by the facilitator.

Performance Assessment. Types of measures, development of instruments, and standard setting

As with testing procedures, intended learning outcomes for each educational goal provide the criteria for judging the quality of the performance. There are two basic types of performance assessment procedures:

➢ *Analytical Scoring.* Those who view student performance as a set of distinct skills rather than as a global generalized ability tend to prefer analytic scoring methods, based on the notion that individual students may have strengths in some areas and deficiencies in others. This procedure uses a recording device such as a rating scale or checklist. The traits of good performance are broken down into categories. Within each category the rater makes a judgment regarding how performance fares on each of the particular dimensions. Each subscale is usually accompanied by a rubric containing detailed descriptors of the characteristics of performances meriting a particular score. Scores on the subscales are then typically added to derive a total score. The main advantage of this procedure is that it provides a *diagnostic* picture of a student's strengths and weaknesses. The main disadvantage is that analytical scoring involves significant amount of time and effort.

➢ *(Modified)Holistic Scoring*. The primary assumption underlying holistic scoring is that the whole performance is more than sum of its parts. This procedure classifies the *overall* performance into one of a number of categories (e.g., Excellent, Good, Inadequate). Each category may be then briefly described by degrees to which the overall criteria have been satisfied. These scoring guides are called scoring rubrics. The main advantage of holistic scoring is speed, rendering it more practical than the analytical scoring approach. The disadvantage is the lack of diagnostic information about students' strengths and weaknesses and compromised reliability (heavy reliance on subjective judgment).

Rating Scales

The rating scale provides a convenient observational guide for analytical rating. It
- focuses attention on the specific elements of the performance to be observed, and
- provides a convenient method of recording the judgments of the observer.

The rating scale can be prepared by using specific learning outcomes as criteria for judging the performance and adding a scale for rating on each specific method. A table below illustrates a rating scale for eight specific learning outcomes comprising general educational goal "Uses laboratory equipment properly"

Characteristics	Rating				
	1	2	3	4	5
Selects appropriate equipment for an experiment					
Assembles equipment correctly for the experiment					
Manipulates equipment as needed					
Measures accurately with proper measuring device					
Follows safety rules					
Uses materials without wasting any					
Completes experiment on time					
Cleans equipment and returns to proper place					

Rating scale: 1=very poor; 2=flawed; 3=average; 4=competent; 5=excellent.

Scoring Rubrics
Characteristics of Scoring Rubrics

- Scoring rubrics are derived from the intended performance outcomes.
- Holistic scoring looks at the overall effectiveness of the student's performance product, without judging individual traits. Performance is viewed as a unified, coherent whole.
- In holistic assessment, the standard is set within the group of papers being assessed, rather than imposing an "ideal" standard.
- Holistic scoring requires evaluations by several faculty members.

- The basis of holistic scoring is a quick and immediate judgment on each performance product. For example, individual papers should be read in one to two minutes maximum, usually in folders of 20 papers each. Thus, readers should read one folder in about 40 minutes, taking at least a 10-minute break before reading the next folder.
- Before scoring begins, faculty score practice papers or other performance products – "range finders" – and discuss standards for scoring.
- Readers may disagree by one point in a judgment on a paper. If the disparity is two points or more, a third reader (or table leader) will evaluate student's product and make a decision.
- Holistic scoring emphasizes what is right rather than what is wrong with student's product
- All judgments are independent; readers should not see one another's scores.
- Holistic scoring is not the best way to evaluate all student works. It offers no feedback to students.
- When student papers will be assessed holistically, they should have access to the scoring rubric before they take exam. Discussing rubrics and using them for classroom assessment can help students develop standards for effective performance.
- When standards are carefully established and raters are carefully trained, holistic scoring produces high reliability.
- A holistic scoring session can promote communication about the teaching among faculty members.

Developing Scoring Rubrics.

There are two basic methods for developing scoring rubrics – a-priori criteria and no a-priori criteria.

Method A: A-priori criteria
1. Have every department member (or potential faculty referee) list 5 to 10 criteria by which a paper should be judged.

2. List all *N* of these on a blackboard and have all referees independently rank-order them from most important (1) to least important (*N*).

3. Sum the ranks assigned to each criterion.

4. List the 10 (or 5 or 12 or whatever number captures the most essential features) criteria with the lowest sum of ranks. Treat these as a *provisional* list of judgmental criteria.

5. Distribute a collection of specimen work products[4] and have each referee independently judge each specimen as satisfactory or unsatisfactory. Each referee should note the criterion or criteria (from the provisional list) that most influenced his or her judgment. If the judgment involved some property that is not on the list, this too should be noted.

6. Discard all products for which there was unanimous judgment (i.e., those that all referees judged satisfactory and those that all referees judged unsatisfactory)

---

[4] If these are materials collected in earlier courses, make sure they represent the full range of quality, from outstanding to unacceptable, and be sure that grades and comments do not appear.

7. Discuss the judgmental criteria noted for those products on which there was inter judge *disagreement*, using the disputed work products to illustrate the properties that exemplify those criteria as understood by the referees. This step is the most important, because it will identify criteria that the judges interpret differently, criteria on which judges have different thresholds of acceptability, and criteria that were omitted from the provisional list but which emerge as important in practice.

8. On the basis of (7), try to reach consensus on how the criteria should be defined and appropriate standards of acceptability for each criterion. If judges consistently attach the *same* meaning to two criteria, one or the other can be deleted. If judges attach *different* meanings to the *same* criterion, criteria can be added to the list to capture the distinctions. Criteria that remain intractably susceptible to diverse interpretation should be discarded.

9. Commit the definitions and examples of acceptability to writing, to be included as part of the instructions for using the criteria.

10. Repeat step (5) using a new set of specimen work products. If there is satisfactory inter judge agreement (e.g., 90 percent), you are finished. If substantial disagreement remains, repeat steps (6), (7), and (8).

Conduct at least one or two field tests. That is, perform step (5) on at least the first set of real data collected for assessment purposes.

Method B: No a-priori criteria
1. Distribute a collection of specimen work products and have each referee independently judge each specimen as satisfactory or unsatisfactory. Each referee should note the features or properties of the paper that shaped his or her judgment. Judgments ordinarily turn on surprisingly few criteria--seldom more than 10, often as few as 3.

2. Discard all products for which there was unanimous judgment (i.e., those that all referees judged satisfactory and those that all referees judged unsatisfactory).

3. For those products on which there was inter judge *disagreement*, list the criteria noted by all judges.

4. Clarify and try to reach consensus on the meaning of the criteria noted by the referees-- how the criteria should be defined, appropriate standards of acceptability for each criterion, etc. Referees should use the disputed work products to illustrate the properties that exemplify the criteria they noted. Such discussion often reveals that different judges attach the *same meaning* to *different words*, and criteria can therefore be deleted from the list. Conversely, judges often attach very *different meanings* to the *same word*, in which case criteria can be added to capture the distinctions. Criteria that remain intractably susceptible to diverse interpretation should be discarded.

5. List all *N* of these criteria on a blackboard and have all referees independently rank-order them from *most important (1)* to *least important (N)*.

6. Sum the ranks assigned to each criterion.

7. List the 10 (or 5 or 12 or whatever number captures the most essential features) criteria with the lowest sum of ranks.

8. Commit the explanation of the criteria (e.g., definitions and examples of acceptability) to writing, to be included as part of the instructions for using the criteria.

9. Test these evaluative criteria by repeating step (1) with a new set of specimen work products. If there is satisfactory inter judge agreement (e.g., 90 percent), you are finished. If substantial disagreement remains, repeat steps (2), (3), and (4).

Conduct at least one or two field tests. That is, perform step (1) on at least the first set of real data collected for assessment purposes.

A relatively simple sample-scoring rubric for assessing critical thinking skills is presented below.

The Critical Thinking Rubric

1) Identifies and summarizes the *problem/question* at issue (and/or the source's position).	
Scant	Substantially Developed
Does not identify and summarize the problem, is confused or identifies a different and inappropriate problem. Does not identify or is confused by the issue, or represents the issue inaccurately.	Identifies the main problem and subsidiary, embedded, or implicit aspects of the problem, and identifies them clearly, addressing their relationships to each other. Identifies not only the basics of the issue, but recognizes nuances of the issue.

2) Identifies and presents the student's own *perspective and position* as it is important to the analysis of the issue.	
Scant	Substantially Developed
Addresses a single source or view of the argument and fails to clarify the established or presented position relative to one's own. Fails to establish other critical distinctions.	Identifies, appropriately, one's own position on the issue, drawing support from experience, and information not available from assigned sources.

3) Identifies and considers other salient *perspectives and positions* that are important to the analysis of the issue.	
Scant	Substantially Developed
Deals only with a single perspective and fails to discuss other possible perspectives, especially those salient to the issue.	Addresses perspectives noted previously, and additional diverse perspectives drawn from outside information.

4) Identifies and assesses the key *assumptions*.	
Scant	Substantially Developed
Does not surface the assumptions and ethical issues that underlie the issue, or does so superficially.	Identifies and questions the validity of the assumptions and addresses the ethical dimensions that underlie the issue.

5) Identifies and assesses the quality of *supporting data/evidence* and provides additional data/evidence related to the issue.

Scant	Substantially Developed
Merely repeats information provided, taking it as truth, or denies evidence without adequate justification. Confuses associations and correlations with cause and effect. Does not distinguish between fact, opinion, and value judgments.	Examines the evidence and source of evidence; questions its accuracy, precision, relevance, completeness. Observes cause and effect and addresses existing or potential consequences. Clearly distinguishes between fact, opinion, & acknowledges value judgments.

6) Identifies and considers the influence of the *context* on the issue.

Scant	Substantially Developed
Discusses the problem only in egocentric or sociocentric terms. Does not present the problem as having connections to other contexts-cultural, political, etc.	Analyzes the issue with a clear sense of scope and context, including an assessment of the audience of the analysis. Considers other pertinent contexts.

7) Identifies and assesses *conclusions, implications and consequences*.

Scant	Substantially Developed
Fails to identify conclusions, implications, and consequences of the issue or the key relationships between the other elements of the problem, such as context, implications, assumptions, or data and evidence.	Identifies and discusses conclusions, implications, and consequences considering context, assumptions, data, and evidence. Objectively reflects upon their own assertions.

PLANNING ASSESSMENT: ADMINISTRATIVE DESIGN

Administrative design is guided by three fundamental questions

- ➢ How many participants should be included?
- ➢ How should they be selected?
- ➢ How often and when should the data be collected?

Sample Size

To answer the question "How many participants should be included?" we have to answer two related questions
- What is the purpose of the study? What questions are being asked?
- How large are effects likely to be? What statistical power will suffice for our purposes?

If the purpose of the assessment is to provide feedback to students about their performance, then the assessment is usually administered to all students. If the purpose of the assessment is to provide the department or institution with information about the performance of students in general, sampling is more appropriate.

Power refers to the likelihood that a difference will be found to be statistically significant for a given sample size. For example, if the purpose of the assessment is to examine the differences in the scores of freshmen and seniors, samples of as few as 20 in each group will produce an 80% chance of finding statistically significant differences between groups. If the purpose of the assessment is to examine the difference between the students at the beginning and end of the freshman year, a sample of 60 to 70 students in each group will be needed. If the purpose of assessment is to examine differences in the gains of two groups over time, 500 students in each group are needed.

There are number of online calculators of sample size. For example, UCLA statistics website (http://calculators.stat.ucla.edu/powercalc/) offers a calculator that computes power for given sample size, or sample size for given power, in a large number of one-sample and two-sample situations. Additionally, a number of other useful calculators allowing performing basic statistical analyses online can be found at http://calculators.stat.ucla.edu/ .

The table below helps to determine required sample size as a function of population size, desired accuracy, and level of confidence.

	Confidence Level	95%	90%	85%	95%	90%
Size of the Population	Sampling Error	5%	5%	5%	1%	1%
50		44	42	40	50	50
100		79	73	67	99	99
200		132	116	102	196	194
500		217	178	147	476	466
1,000		278	216	172	907	873
2,000		322	242	188	1,661	1,550

For example, if you are sampling from a universe of fifty people and you want to have 90% confidence that your results will be within 5 % of the true percentage in the population, you need to randomly sample forty-two people.

Sample Selection

In general, the sample should be representative of the population from which the sample was drawn to allow making generalizations about the population. The best way of assuring that a sample is representative is to use random selection of students (e.g., by using SPSS random numbers procedure.)

Stratified random sampling is often used when faculty members want to make sure they have adequate numbers of students in some subgroups. The overall group of interest is divided into subgroups based on a categorical variable such as class standing, declared major, classes taken, race, gender, etc. Individuals are then randomly selected within the subgroups

In qualitative research, a number of case selection methods have been suggested.

- *Extreme cases*. The cases demonstrate unusual manifestation of the phenomenon, such as outstanding success and notable failures.
- *Intensity cases*. The cases are information rich but not extreme cases.
- *Maximum variation*. Cases, despite having diverse variations, exhibit important common patterns that cut across variations.
- *Homogeneous.* Variation between cases is minimized, analysis is simplified and study is focused.
- *Typical case.* Case illustrates what is typical, normal or average.
- *Stratified purposeful cases*. Cases illustrate characteristics of a particular subgroup to facilitate comparison and not for generalization or representation.
- *Critical cases.* Cases that permit logical generalization to other cases because if it is true to this one case, it's likely to be true to all other cases
- *Snowball.* Cases of interest from people who know people who know people who know cases, rich information, rich, good examples for study, etc.
- *Criterion*. Cases picked because they meet some predetermined criterion.
- *Theoretical.* The cases are manifestation of a theoretical construct and are used to examine and elaborate on it.
- *Confirming and disconfirming.* Cases that elaborate on initial analysis to seek exceptions or test variations.
- *Opportunistic.* Cases that emerge from following leads during fieldwork.
- *Random purposeful.* Cases are randomly selected from a large sample for the purpose of increasing credibility and not for generalization or representation.
- *Politically important cases*. Cases are selected or eliminated because they are politically sensitive cases.
- *Convenience*. Cases are selected on the basis of minimum effort, time and money. They are candidate examples of low credibility, information rich cases.
- *Combination*. Cases are flexible and meet different interests and needs

Mapping Multiple Sources

Mapping multiple sources is about gathering those data that represent different components of the question(s). This may require gathering data from different sources.

- Simultaneous mapping – to conduct assessment with different groups without using the results from prior group analyses
- Sequential or linear mapping – to conduct assessments with different groups in an order that allows the use of results from one group to the next
- Circular mapping -- to conduct assessments with different groups using the results from one group to the next, then revisiting a particular group for further assessment

Timeline

The answer to the question "How often should the data be collected?" depends on the purpose of the assessment and the chosen method. Data collection, for example, may occur at the end of a semester in a capstone course, involve a one-time project, or occur on a yearly basis. For many program goals, it may be useful to assess student understanding when they enter the major. An entry level required course that serves as an introduction to the major can serve as a fine pre-test data collection point. Both external and internal needs should be considered when the timeline is developed. The schedule of reviews by regional and/or professional agencies and state reporting requirements generally have a major impact on the time line, as does an institutional schedule for program review.

Different groups of students may need to be assessed at different times. For example, evaluating graduating seniors near the end of their last semester is clearly a critical time. Evaluation of alumni can happen at any time.

Dates and policies for assessment endeavors should be set at the start of the academic year and distributed to students and faculty.

Successful assessment plans are driven by three calendars:

1. One that develops assessments *for each goal* separately. For example, one successful assessment plan with separate goals might put the development or review of assessments on a phased calendar by year.

Goals	Year 1	Year 2	Year 3
1.	X		
2.	X		
3.		X	
4.		X	
5.			X
6.			X

2. One that describes the year-to-year operation of the assessments (*assessment cycle*). For example, a successful assessment plan might use an academic calendar with tasks identified for each month or semester.

Sample Departmental Assessment Cycle Timeline

	Fall Semester (beginning)	*Fall Semester (end)*	*Winter Break*	*Spring Semester (beginning)*	*Spring Semester (End)*	*Summer Break*
Year One						
Development	Department discussions regarding Mission, Goals, Learning Outcomes	>>>>>>>>	Complete Statements of General Goals and Specific Learning Outcomes	Develop Assessment Strategies and three year plan	>>>>>>>>	>>>>>>>
Year Two						
Data Collection		*Sophomores focus groups *Course embedded competency testing		Capstone Assignment	*Senior survey *Course embedded competency testing	
Analysis			*Focus Group analysis *Competency analysis		*Department scores capstone *Competency analysis	
Reporting/ Use					*Competency reports *Report cards	
Year Three						
Data collection						
Analysis						
Reporting / Use						

3. One that develops assessment *for each specific outcome/ assessment strategy* separately. For example, one successful assessment plan with separate outcomes might put the development or review of assessments on a calendar by month.

Sample departmental assessment calendar

Outcome	Assessment Task/Strategy	Month Attained					
		Aug.*	Dec.*	Feb.*	Mar.*	Apr.*	May**
Maintain a high level (80% or higher) of overall employer	Conduct focus groups					X	X
	Analyze data					X	X

---

* Example – Please adjust as appropriate.

Outcome	Assessment Task/Strategy	Month Attained					
		Aug.*	Dec.*	Feb.*	Mar.*	Apr.*	May*[*]
overall employer satisfaction with the preparation of program graduates for entry-level work at the bachelor's degree level.	Provide feedback to faculty	X					
After completing the general education writing sequence and completing college-level courses into the junior year, students should possess the characteristics found in competent writers as defined in the University's Writing Competency Assessment Plan.	Collect writing portfolios from juniors (prior to earning 90 credits)		X			X	
	Evaluate writing portfolios; provide feedback to students; return portfolios to students		X				X
	Provide feedback to faculty	X			X		
After completing a departmental research methods course, students should know the major research designs, the basic steps in planning and conducting research, and be able to apply an appropriate research strategy to address a significant problem in the field as demonstrated in a research methods term paper. Assessment Strategies 1. ETS Field Test 2. Assess Research Methods' Course Term Papers	1a. Administer standardized test to seniors in XYZ major (e.g., ETS Field Test in Psychology)		X			X	
	1b. Analyze research methods content data from examination (research methods' content scores and group performance on individual items related to research methods)			X			X
	2a. Select a sample of research methods term papers		X				X
	2b. Evaluate term papers in terms of research competencies identified by program faculty as important for graduates to know and be able to do		X				X
	Provide feedback to faculty	X			X		

Attachment 3 provides more examples of assessment timing and participant selection in a variety of academic programs.

Once you have established a planning calendar, you might want to identify fixed and variable costs associated with each step in the assessment process.

Task	Fixed Costs	Variable Costs
Set or review purpose(s)	2 hr x 1 mtg x 6 member committee	
Set or review general educational goals	2 hr x 3 mtg x 6 member committee	Survey of 100 selected faculty
Identify specific learning outcomes that reflect general goals	2 hr x 3 mtg x 6 member committee	Purchase sample measures
Determine criteria to measure outcomes		Costs of invitations, tests, survey, follow-up
Select multiple measures		
Formulate implementation plan		
Select and invite pilot student group		
Administer assessment		
Analyze and evaluate data		
Program reform		
Professional development		
Public reporting		

Template for Program Assessment Plan

A typical assessment plan is built around the following topics:

1. *Program goals*: Describe what program intends to accomplish, how the program's goals relate to institutional mission and goals, general education goals, and/or state-mandated core competencies.
2. *Specific learning outcomes*: Describe what students must know and can do.
3. *Methods:* Describe what methods will be utilized to assess the outcomes. For example, multiple-choice test, portfolio, essay, presentation, survey.
4. *Instrument:* Describe the instrument used for assessment. For example: commercially available test (in attachment, provide reliability[5] and validity[6] information about the test); in-house test (in attachment, provide test blueprint and reliability coefficients if available); portfolio (in attachment, describe structure – components and scoring rubrics); essay (in attachment, describe scoring rubrics). Indicate cut-off scores / standards (in attachment, describe how they were determined.)
5. *Criteria for success*: Indicate how you will determine whether learning objectives have been met. Answer the question "If our instructional program is functioning the way we think it 'ought' to function, what proportion of our students achieve the level of competency on the given outcome?". For example, "two-thirds of students pass a scientific reasoning test at the first attempt" or "90% of program graduates are employed during the first year after graduation."
6. *Target groups:* Describe the population to be assessed and the sample (size, method of selection)
7. *Time line*: Indicate when data will be collected and analyzed,

---

[5] Reliability coefficients if available
[6] Analysis of test content (matching items with program objectives), if available.

8. *Provisions for administration*: Describe where assessment is located in program activities / curriculum (e.g., course or program requirement, general education or graduation requirement, etc.) Indicate where assessment will take place (e.g., classroom, lab.) Indicate who has responsibility for seeing the plan is carried out, who will collect and analyze data, and who will report results. Indicate resources (training) needed and estimate costs.

9. *Reporting*: Describe when reports will be available, when recommendations will be made, provisions for sharing information with internal and external audiences, and for making recommendations and decisions.

The following template may assist in developing program assessment plan.

FORMAT FOR ACADEMIC PROGRAM ASSESSMENT PLANS

PROGRAM MISSION
I.1. Brief program overview (structure, courses, staffing)
I.2. Program mission (with the reference to the university mission)

PROGRAM GOALS
II.1. Goal 1 (with the reference to the university goals)
II.2. Goal 2 (with the reference to the university goals)
II.3. Goal 3 (with the reference to the university goals)

INTENDED STUDENT LEARNING OUTCOMES[7]
III.1. Intended earning outcome(s) for Goal 1
    III.1.1. Outcome 1/1
- Where in the *curriculum* the outcome is addressed? [i.e., where in the curriculum the outcome is developed and where the attainment of the outcome is measured[8].]
- The *criteria* (standards) for success [i.e., what level of achievement is considered acceptable for students in the program.]
- At least 2 (two) assessment *method*s used to determine whether the outcome was attained [including how, when, and by whom attainment of the outcome was measured.]
III.2. Intended learning outcome(s) for Goal 2
III.3. Intended learning outcome(s) for Goal 3

PROGRAM OUTPUTS
IV.1. Output 1
- The *criteria* (standards) for success [i.e., what level of performance is considered acceptable for the program]
- The assessment *method*(s) used to measure whether the output was attained [including how, when, and by whom attainment of the outcome was measured]

---

[7] **It is advisable to limit the number of learning outcomes to no more than one or two for each goal.**
[8] See *Assessment Manual,* p.20 (Curriculum Audit Grid)

IV.2. Output 2
IV.3. Output 3

ESTIMATED COSTS OF ASSESSMENT (OPTIONAL)

SUMMARY

Plan Completed by: _____ **Date:** _____

Department Chair: _____ **Date:** _____
(if different from person completing the report)

REVIEW CRITERIA FOR ACADEMIC PROGRAM ASSESSMENT PLANS

MISSION

- ❑ Concise
- ❑ Lists stakeholders
- ❑ States purpose / means to accomplish the purpose / values
- ❑ Supports NSU's and School's missions
- ❑ Uniquely related to the program

PROGRAM GOALS

- ❑ Related to the mission
- ❑ Reflect state mandates and institutional expectations about learning (e.g., the six core competencies[9])
- ❑ Cover the wide range of content topics, skills, attitudes expected in the program
- ❑ Attainable (do students have the necessary experiences and educational background to proceed successfully?)
- ❑ Reflect student interests
- ❑ Retention (do these goals reflect learning outcomes that tend to be retained longest, e.g., comprehension, application, thinking skills?)
- ❑ Transfer value (do these goals reflect learning outcomes that are widely applicable to new situations? Do the goals reflect realistic and complex learning tasks that are most useful in the "real world"?)

---

[9] Written Communication, Technology/Information Literacy, Scientific Reasoning, Quantitative Reasoning, Critical Thinking, and Oral Communication

STUDENT LEARNING OUTCOMES

- ❑ Student-focused rather instructor-centered
- ❑ Focus on the learning resulting from an activity rather than the activity itself
- ❑ Focus on important, non trivial aspects of learning that are credible to the public
- ❑ Measurable
- ❑ Are effectively worded

- ▪ Use action verbs that describe definite, observable actions
- ▪ Include a description under which the action takes place – "when given X, the student will be able to…"
- ▪ Indicate an appropriate level of competency (standard) that is assessable through one or more indicators

ASSESSMENT METHODS

- ❑ Justification of the selected method(s)
- ❑ Manageability
- ❑ Assessment instruments (rubrics, grades criteria, survey, test, etc.) are attached
- ❑ Was an externally created instrument considered? Was an in-house instrument considered? If an in-house instrument was adopted, how were the test items constructed?
- ❑ Evidence of reliability and validity
- ❑ Standard setting procedures are clearly described and appropriate
- ❑ Sampling methods are clearly described and appropriate
- ❑ Assessment administration procedures are clearly described

Gaining Cooperation from Students

Useful information can be collected only if students take the assessment activities seriously. The following guidelines can help faculty members gain student cooperation.

> ➢ Do not surprise students. Early and repeated announcements should be made about any required assessment activity, including its purpose and the students' role.

> ➢ Explain the purpose and expectations of assessment. A good practice is to include assessment requirements in the program brochure (or institutional catalogue) in the form of a statement that carefully lists the times for assessment and explains how the results will affect students' academic standing. For example:

> > Student Outcome Assessment Consent Form
> >
> > In order to comply with the guidelines of the State Council of Higher Education in Virginia (SCHEV), the Southern Association of Colleges and Schools (SACS), and *** Program Accrediting Agency, *** Program requires a series of student outcome assessments. I understand that these assessments are required part of my education experience and will be kept confidential. Assessment results will be used along with other students' scores to identify relative strengths and weaknesses in the educational program, so that ***Program can continually improve. I agree to participate in such assessments when requested over the course of my undergraduate period.

> ➢ Include students on assessment planning committees and use existing student groups for publicity.
>   - Placing at least one student on each committee will ensure student input and create another link between students and faculty
>   - To generate interest and focus attention on assessment, articles can be featured in program newsletter and student publications just prior to a major assessment activity
>   - Conduct an information session on assessment for the members of program student association / student government

> ➢ Explain what's in it for students. Some students do not visualize a direct benefit of assessment for their education unless they receive individual feedback. Programs that focus only on summarizing information might add another component: reporting assessment results to individual students for their own learning and development.

COMMUNICATING RESULTS EFFECTIVELY

General Guidelines

General Principles for Communicating Assessment Results Effectively

➢ *Integration.*
Data and information should be presented in relation to identified program goals and specific learning outcomes. Develop recommendations based on analysis of data, and use identified goals and outcomes as a framework within which to accomplish these changes.

➢ *Assessment results need to be communicated frequently.*
The number of institutional assessment studies conducted and reported is one of the best predictors of the efficacy of assessment. Continuing communication of assessment results also helps break down the functional silos within institutions replacing them with a sense of community and common purpose.

➢ *Know your audience.*
It is important to know who the decision makers are and to ensure that they receive appropriate information upon which they base their decisions. It is also important to know what types of information the decision makers prefer and how they like to have results reported. The effective dissemination of assessment results requires multiple modes of communication that are tailored to the needs of a given audience.

➢ *Understand your information.*
An important consideration in deciding how to communicate assessment results is understanding which presentation format most clearly communicates the findings of the study. Use qualitative and quantitative methods to present a well-balanced holographic and dynamic picture of the program.

Also consider the extent to which your findings can help you answer the following questions.

- What do the data say about your students' mastery of subject matter, of research skills, or of writing and speaking?
- What do the data say about your students' preparation for taking the next step in their careers?
- Are there areas where your students are outstanding? Are they consistently weak in some respects?
- Are graduates of your program getting good jobs, accepted into reputable graduate schools, reporting satisfaction with their undergraduate education?
- Do you see indications in student performance that point to weakness in any particular skills, such as research, writing, or critical thinking skills?
- Do you see areas where performance is okay, but not outstanding, and where you would like to see a higher level of performance?

Analyzing Assessment Information

Descriptive and Comparative Information

Descriptive information includes responses to various questions on surveys, mean scores on assessment examinations, and summaries of scores assigned to various products as well as information about individuals who participated in the study, including the number of cases and some indication how well they represent the overall group of interest on key characteristics.

Because student learning vary, it is a useful approach to compare results for various groups of students. *Comparative* information can be obtained by grouping students in several ways:
- demographic characteristics (e.g., gender, race, on-campus/off-campus residence)
- educational background (e.g., SAT scores, high school rank)
- college experiences (e.g., test results can be compared for freshmen and juniors, for the students who have taken a specific course and those students who have not)

Impact of Various Response Scales on Analysis

Certain response scales limit the types of statistics that can be calculated.

- Responses that are expressed in *categories* such as gender or class standing can be reported in tables showing the numbers or percentage of cases in each group, but do not allow calculating means or medians. Cross-classification tables can be used to display results in categories. Content analysis applied to open-ended questions on surveys or comments from focus groups represents one way to generate categorical data. Comments are coded and sorted into various categories.
- Both medians and modes (but usually not means) can be calculated for response scales that capture *ranks* or orders, such as "strongly agree" to "strongly disagree." Scoring scales used for performance assessment represent rank data
- Scaled or *interval* data such as age, height and scores on multiple-choice and other objective tests are expressed in meaningful numbers. These data allow several kinds of analyses, including descriptive statistics (e.g., means, medians, modes) and multivariate analysis[10] (e.g., ANOVA, regression)

---

[10] Advanced multivariate analyses can also be performed with interval and rank order data (e.g., logistic regression).

Analyzing the Reliability of Assessment Data

To demonstrate *credibility* of assessment data, it is a good practice to report the following indices.

1. The test item difficulty index.

The test item difficulty index simply represents the proportion or percentage of students answering the particular test item correctly. For a norm-referenced interpretation, test item difficulties should be around .5. For a criterion-referenced framework, the item difficulties are usually higher than .5, depending on expected competency levels.

2. The item-score correlation index.

This index is the correlation between getting a particular test item correctly and the total item score. High values for a question indicate that students who answered that question correctly are also scoring high on the test itself.

3. The item discrimination index.

When used with norm-referenced interpretations, it is a measure of the difference between the proportion of students with high test scores answering the test item correctly and the proportion of students with low test scores answering the same item correctly. High and low test scores are usually defined as the top 27% and the bottom 27% of the sample. Questions with negative discrimination differences may be problem questions, since low-scoring students were more likely to answer the test item correctly than high scoring students.

When used with criterion-referenced interpretations, this index examines "instructional sensitivity". Instructional sensitivity measure uses competency or cutoff scores: $B = (U/n1) - (L/n2)$, where $n1$ is the number of students scoring along the cutoff level, $n2$ is the number scoring below, U is the number above the cutoff who answered the particular item correctly, and L is the number of students below the cutoff who answered the item correctly. B ranges from -1.0 to +1.0, with higher positive values sought for each item.

Analytical Strategies for Processing Assessment Information

- Do students change over time? Typically at least two comparison groups of students are used: one participating in the educational program or course and the other control group who has not participated. Do students who participated in the experience change to a greater extent than students who did not participate?
- If a pretest measure is not available, do students who participate in the educational experience score or rate higher than students who did not participate?
- What is the relationship or correlation of the outcome measure compared to other measures of student performance or success? For example, are class grades related to

competency testing scores? Are program outcomes measures related to measures of later job performance or citizenship?

- Are students performing at expected levels of competency or how many students reach your academic standard after program completion? What differences are there between students who pass and those who do not pass?

Using Assessment Information

When the faculty makes use of assessment information, they can consider a large number of program facets:

➢ *Program processes*
- advising procedures
- use of pretest for admittance to courses
- integration of materials across courses
- course offerings
- out-of-class support
- suggested/required order for taking courses
- program policies (grading, attendance, etc.)
- faculty course assignments
- administrative policies

➢ *Program inputs*
- number of faculty
- faculty training and expertise
- upgraded facilities and equipment
- budget allocations
- student requirements for high school coursework

➢ *Assessment processes*
- choice of methods
- standard setting
- timing of assessment
- sampling procedures
- use of inducements for student participation
- faculty responsibilities for process

➢ *Program foundations*
- mission statement
- goals

Changes are not going to be perfect and improvements can be expected to take years. Assessment process helps programs build on strengths and minimize weaknesses

incrementally over time. Assessment processes take three or four years to implement fully and may take even longer to result in drastic program improvement. The strength of assessment is not that it provides quick fixes for any problems we might have, but that it promotes active, informed, and systematic improvement of all aspects of a program.

Reporting

Formal or Comprehensive Report

The formal report for the program typically should include the following sections and reflect the following criteria:

1. Comprehensiveness and specificity of the program's general educational goals and intended learning outcomes.
- Do the goals appear to cover the wide range of content topics expected in the program?
- Do goals include goals for writing? Technology skills? Information seeking skills? Quantitative reasoning? Scientific reasoning? Critical thinking? Oral communication? Leadership skills?
- Are the goals translated in specific measurable outcomes?

2. Assessment methods.
- How appropriate are the assessment methods for evaluating each program goal?
- Was an externally created instrument considered?
- Was an in-house instrument considered?
- If an in-house instrument was adopted, how were the test items constructed?
- Is there evidence of reliability?
- Is there evidence of validity (e.g., table of specifications)?
- How were the standards/criteria established?
- Were the sample selection and assessment administration procedures described?

3. Analyses of results.
- Was the information summarized in a quantifiable form?
- Were the narrative explanations of the meaning of the information provided?
- Is there a discussion of how well the assessment information meets the stated goals and intended outcomes?
- Has the program identified program strengths and weaknesses?
- Why are some students scoring higher or lower?

4. Uses.
- Has the program specified how the information obtained from assessment will be or was used?
- What program changes have been or will be made?

- What other recommendations based on the assessment information have been given?

5. Future assessment plans and goals.
    - What questions will be addressed in future studies?
    - What goals will be covered in future reports that were not covered in this report?

6. Additional information.
    - Executive summary
    - Cost of assessment activities
    - Appendix containing a curriculum analysis matrix, relevant assignments and outcomes, data collection methods, and other information or materials as appropriate

A FORMAT FOR ACADEMIC PROGRAM ASSESSMENT REPORTS

PROGRAM MISSION

I.1. Brief program overview (structure, courses, staffing)
I.2. Program mission (with the reference to the university mission)

PROGRAM GOALS

II.1. Goal 1 (with the reference to the university goals)
II.2. Goal 2 (with the reference to the university goals)
II.3. Goal 3 (with the reference to the university goals)

STUDENT LEARNING OUTCOMES[11]

III.1. Learning outcome(s) for Goal 1
    III.1.1. Outcome 1/1
- Where in the *curriculum* the outcome is addressed? [i.e., where in the curriculum the outcome is developed and where the attainment of the outcome is measured[12]]
- The *criteria* (standards) for success [i.e., what level of achievement is considered acceptable for students in the program]
- At least 2 (two) assessment *method*s used to determine whether the outcome was attained [including how, when, and by whom attainment of the outcome was measured]
- The *results* of the assessment [i.e., what was learned about how well the program is achieving the outcome]
- *Use* of assessment results for program improvement [i.e., indicate how assessment results are being (will be) used to improve student performance in the specific area)
  - *Changes* to assessment process. For example,
    - Restructure goal/outcomes statement
    - Revise method of data collection/measurement approach
    - Conduct deeper analysis
  - *Changes* to operation or academic process. For example,
    - Revise admission criteria
    - Revise advising standards or processes
    - Streamline program offerings
    - Make technology related improvements
  - *Changes* to curriculum. For example,
    - Revise prerequisites and/or sequences of courses
    - Revise course content
    - Add/delete course(s)

---

[11] It is advisable to limit the number of learning outcomes to no more than one or two for each goal.
[12] See *Assessment Manual,* p.20 (Curriculum Audit Grid)

III.2. Learning outcome(s) for Goal 2

III.3. Learning outcome(s) for Goal 3

PROGRAM OUTPUTS

IV.1. Output 1
- The *criteria* (standards) for success [i.e., what level of performance is considered acceptable for the program]
- The assessment *method*(s) used to measure whether the output was attained [including how, when, and by whom attainment of the outcome was measured]
- The *results* of the assessment [i.e., what was learned about how well the program is achieving its objective regarding the output measure]
- *Use* of assessment results for program improvement [i.e., indicate how assessment results are being (will be) used to used to improve student performance in the specific area)
  - *Changes* to assessment process. For example,
    - Restructure goal/outcomes statement
    - Revise method of data collection/measurement approach
    - Conduct deeper analysis
  - *Changes* to operation or academic process. For example,
    - Revise admission criteria
    - Revise advising standards or processes
    - Streamline program offerings
    - Make technology related improvements

  - *Changes* to curriculum. For example,
    - Revise prerequisites and/or sequences of courses
    - Revise course content
    - Add/delete course(s)

IV.2. Output 2

IV.3. Output 3

ESTIMATED COSTS OF ASSESSMENT (OPTIONAL)

SUMMARY

BIBLIOGRAPHY and Additional Resources:

Banta, T.W. (Ed.) (2002). *Building a scholarship of assessment*. San Francisco: Jossey-Bass.

Bergen, R. (2000). *A program guide for outcomes assessment at Geneva College*. Geneva College. Available: http://www.geneva.edu/about_geneva/academic/outcome_assessment/oaguide.pdf

Brown, S., & Glasner, A. (Eds.) (1999). *Assessment matters in higher education: Choosing and using diverse approaches*. Philadelphia: SRHE and Open University Press.

Bloom, B. S. (Ed.) (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain.* White Plains, NY: Longman.

Bond, L.A. (1996). Norm- and criterion-referenced testing. *Practical Assessment, Research, & Evaluation, 5* (2). Available online: http://ericae.net/pare/getvn.asp?v=5&n=2

Buckhendahl, C.W., Smith, R.W., Impara, J.C., & Plake, B.S. (2001, April). *A comparison of Angoff and Bookmark standard setting methods.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, Seattle, WA.

Committee on Assessment, Accreditation, and Strategic Planning (1999). *Assessment handbook*. University of Michigan-Flint.

Concordia College (Online). *Concordia College assessment handbook*. Available: http://www.cord.edu/dept/assessment/ahbcontents.html

DeMars, C.E., Sundre, D.L., & Wise, S.L. (2002). Standard setting: A systematic approach to interpreting student learning. *JGE: The Journal of General Education, 51* (1), 1-19.

Erwin, T.D. (2000). *The NPEC sourcebook on assessment*, Volumes 1 and 2. U.S. Department of Education, NCES. Washington, DC: U.S. Government Printing Office.

Erwin, T.D. (1999). Assessment and evaluation: A systems approach. In: Brown, S., & Glasner, A. (Eds.), Assessment matters in higher education: Choosing and using diverse approaches, pp. 28-40. Philadelphia: SRHE and Open University Press.

Erwin, T.D. (1991). *Assessing student learning and development: A guide to the principles, goals, and methods of determining college outcomes*. San Francisco: Jossey-Bass.

Gronlund, N.E. (2000). *How to write and use instructional objectives*, 6th edition. Upper Saddle River, NJ: Merill / Prentice Hall

Gronlund, N. E. (1981). *Measurement and Evaluation in Teaching*. 4th ed. New York: Macmillan.

Hatfield, S.R. (1999). Department level assessment: Promoting continuous improvement. *IDEA Paper, 35*. Manhattan, KS: IDEA Center, Kansas State University.

Kiplinger, V.L. (1997). *Standard-setting procedures for the specification of performance levels on a standard-based assessment*. Available: http://www.cde.state.co.us/cdeassess/asperf.htm

Nichols, J.O. (1989). *Institutional effectiveness and outcomes assessment implementation on campus: A practitioner's handboo*k. New York: Agathon Press.

North Carolina State University / University Planning and Analysis (2002). *Internet resources for higher education outcomes assessment.* Available: http://www2.acs.ncsu.edu/UPA/assmt/resource.htm

Palomba, C.A., & Banta, T.W. (Eds.) (2001). *Assessing student competence in accredited disciplines: Pioneering approaches to assessment in higher education*. Sterling, VA: Stylus..

Palomba, C., Pickerill, B., Shivaswamy, U., Woosley, S., Moore, D., Shaffer, P., & Stout, T. (2000). *Assessment Workbook*. Ball State University. Available: http://web.bsu.edu/IRAA/AA/WB/contents.htm

Palomba, C.A., & Banta, T.W. (1999). *Assessment essentials: Planning, implementing, and improving assessment in higher education.* San Francisco: Jossey-Bass.

Patton, M.Q. *Qualitative Evaluation and Research Methods*, Sage Publications, Newbury Park, California, 1990.

Pike, G.R. (2002). Measurement issues in outcomes assessment. In: Banta, T.W. (Ed.), *Building a scholarship of assessment*, pp.131-147. San Francisco: Jossey-Bass.

Shupe, D.A. (2001). Reassessing assessment. *Assessment Update, 13* (5), 6-7.

Stassen, M.L.A., Doherty, K, & Poe, M. (2001). *Program-based review and assessment: Tools and techniques for program improvement.* University of Massachusetts. Available: http://www.umass.edu/oapa/assessment/program_based.pdf

St.Cloud State University / Assessment Office (Online). *Guidelines for program assessment.* Available: http://condor.stcloudstate.edu/~assess/guidelines.html

Student Outcomes Assessment Committee (Online). *Assessment: An institution-wide process to improve and support student learning*. College of DuPage. Available: http://www.cod.edu/outcomes

Troy, M., & Lowe, P.A. (Online). *Texas A&M University assessment manual*. Available: http://www.tamu.edu/marshome/assess/Manual.html

Twelker, P.A. (1999). *Guidelines for developing departmental assessment plans. Trinity International* University. Available: http://www.tiu.edu/psychology/Guidelines.htm

University of Wisconsin-Madison (2000).*UW-Madison assessment manual: Using assessment for academic program improvement.* Available: http://www.wisc.edu/provost/assess/manua.html

University Testing Services (2000). *Resources*. Pennsylvania State University. Available: http://www.uts.psu.edu/

VCCS (2002). *Report of the VCCS taskforce on assessing core competencies*. Available: http://www.vccs.edu/competencies/taskforcereport.pdf

Volkwein, J.F. (1998). Managing a program of outcomes assessment. *AIR 37th Annual Forum Pre-Forum Workshop*. Available: http://www.rose-hulman.edu/IRA/IRA/IR/inair/content.html

Washington State University (2002). *CT Project: The critical thinking rubric*. Available: http://wsuctproject.wsu.edu/ctr.htm

Selected Assessment Methods References:

Case studies.
Graziano, A.M., & Raulin, M.L. (1993). *Research methods: A process of inquiry*, 2nd edition. New York: HarperCollins.

Ragin, C.C. (1994). *Constructing social research: The unity and diversity of method*. Thousand Oaks, CA: Pine Forge Press.

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

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

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

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

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

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

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

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

Essay and Objective questions.
Jacobs, L.C., & Chase, C.I. (1992). Developing and using tests effectively: A guide for faculty. San Francisco: Jossey-Bass.
Ory, J., & Ryan, K.E. (1993). Tips for improving testing and grading. Beverly Hills: Sage

Focus groups.
Morgan, D.L., & Krueger, R.A. (1997). *The focus group kit*. Thousand Oaks, CA: Sage.

Interviews.
Rubin, H.J., & Rubin, I.S. (1995). *Qualitative interviewing: The art of hearing data*. Thousand Oaks: Sage.

Matrices.
Hatfield, S., Krueger, D., & Hatfield, T. (1998, June). *Starting points: Developing a department assessment plan.* Paper presented at the AAHE Assessment Conference, Cincinnati.

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

Phone surveys/interviews.
Babbie, E. (1995). *The practice of social research*, 7th edition. Belmont, CA: Wadsworth.

Portfolios.
Black, L.C. (1993). Portfolio assessment. In: Banta, T.W., & Associates (Eds.), *Making a difference: Outcomes of a decade of assessment in higher education*, 139-150. San Francisco: Jossey-Bass.

Reflective Essays.
Schilling, K. (1998, June). *Miami University Liberal Education Project*. Paper presented at the AAHE Assessment Conference, Cincinnati.

Scoring rubrics.
Harris, S. (1998, March). *Using scoring rubrics*. Presentation at CETL Conference on Assessment, Fresno.

Standardized achievement tests.
Allen, M.J. (1995). *Introduction to psychological research*, Itasca, IL: Peacock.

Surveys.
Allen, M.J. (1995). *Introduction to psychological research*, Itasca, IL: Peacock.

Fowler, F.J. (1985). *Survey research methods*. Beverly Hills, CA: Sage.

Transcript analysis.
Hatfield, S., Krueger, D., & Hatfield, T. (1998, June). *Starting points: Developing a department assessment plan.* Paper presented at the AAHE Assessment Conference, Cincinnati.

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

ATTACHMENT 1. Classification of Cognitive Skills

(Bloom, 1956)

Category	Definition	Related Behaviors
Knowledge	recalling or remembering something without necessarily understanding, using, or changing it	define, describe, identify, label, list, match, memorize, point to, recall, select, state
Comprehension	understanding something that has been communicated without necessarily relating it to anything else	alter, account for, annotate, calculate, change, convert, group, explain, generalize, give examples, infer, interpret, paraphrase, predict, review, summarize, translate
Application	using a general concept to solve problems in a particular situation; using learned material in new and concrete situations	apply, adopt, collect, construct, demonstrate, discover, illustrate, interview, make use of, manipulate, relate, show, solve, use
Analysis	breaking something down into its parts; may focus on identification of parts or analysis of relationships between parts, or recognition of organizational principles	analyze, compare, contrast, diagram, differentiate, dissect, distinguish, identify, illustrate, infer, outline, point out, select, separate, sort, subdivide
Synthesis	creating something new by putting parts of different ideas together to make a whole.	Blend, build, change, combine, compile, compose, conceive, create, design, formulate, generate, hypothesize, plan, predict, produce, reorder, revise, tell, write
Evaluation	judging the value of material or methods as they might be applied in a particular situation; judging with the use of definite criteria	accept, appraise, assess, arbitrate, award, choose, conclude, criticize, defend, evaluate, grade, judge, prioritize, recommend, referee, reject, select, support

ATTACHMENT 2. Taxonomy of Learning Areas

(Gronlund, 1981)

Knowledge

- terminology
- specific facts
- concepts and principles
- methods and procedures

Understanding

- concepts and principles
- methods and procedures
- written material, graphs, maps, and numerical data
- problem situations

Application

- factual information
- concepts and principles
- methods and procedures
- problem-solving skills

Thinking skills

- critical thinking
- scientific thinking

General skills

- laboratory skills
- performance skills
- communications skills
- computational skills
- social skills

Attitudes

- social attitudes
- scientific attitudes

Interests

- personal interests
- educational interests
- vocational interests

Taxonomy of Learning Areas (cont'd)

Appreciation

- literature, art, and music
- social and scientific achievements

Adjustments

- social adjustments
- emotional adjustments

ATTACHMENT 3. Selected Assessment Methods / Timeline / Samples for Academic Programs

- *Accounting* - American Institute of CPAs Achievement Test (every semester including all bachelors and masters graduates), Annual Survey of Employers and Recruiters (each year covering all bachelors and masters graduates for which placement data are available), Certified Public Accountant examination (May and November to compare bachelors and masters graduates taking examination with comparable institutions), College Graduating Senior Survey (each year)
- *Administrative Sciences* - Graduate Management Aptitude Test (each semester), Major Field Achievement Test (each semester), College Graduating Senior Survey (each year)
- *Adult, Counseling, Health, and Vocational Education*: *Community Health Education* - Internship (seniors)
- *Adult, Counseling, Health, and Vocational Education*: *Counseling and Human Development Service*s - state counseling licensure test (after graduation), practicum coursework and internship (masters students), dissertation (doctoral students), employment survey (masters and doctoral students)
- *Adult, Counseling, Health, and Vocational Education: School Health Education* - National Teacher Examination (each semester to graduating students), Student Teaching (seniors), National Certified Health Education Specialist Examination (optional examination)

- *Adult, Counseling, Health, and Vocational Education: Rehabilitation Counseling in Education* - National Certification Exam (all masters students), Follow-Up of Employment Status of Graduates (all masters students), Internship Evaluation (all masters students), Oral Examination (all masters students)
- *Adult, Counseling, Health, and Vocational Education: Vocational Education* - National Teacher Examination (each semester to graduating students), Student Teaching (seniors), Oral Examination (all masters students)
- *Anthropology* - Departmental Undergraduate Major Achievement Test (seniors), Senior Survey (yearly), Graduate Student Survey (yearly), Professional Achievement Survey (MA and PhD graduates)
- *Architecture and Environmental Design* - National Council of Architecture Registration Boards examination (each year), National Architectural Accrediting Board evaluation and accreditation (every two to five years, depending on evaluation), Student Design Studio Juries (each semester for each year level)

- *Art: Art Education* - Pre-Professional Skills Test of Reading, Writing, and Mathematics (upon completion of 32 semester credit hours or during first semester after declaring an education major or minor), Art Education Major Review (each semester for seniors prior to acceptance into student teaching), National Teacher Examination (each semester to graduating students)
- *Art: Art History* - Senior Research Project (under consideration)
- *Art: Crafts* - Sophomore Portfolio Review (1 SmHr credit course, required), Senior Project Exhibition (each semester for graduating students)
- *Art: Fine Arts* - Sophomore Portfolio Review (under consideration), Junior Portfolio Review (each semester), Senior Project (each semester)
- *Art: Graphic Design and Illustration* - Entrance Examination (before entry into program), Portfolio Review (spring semester of sophomore year, after completing six introductory courses in GD&I), Portfolio Review (end of junior year), Professional Portfolio (each semester for seniors), Senior Project (each semester for seniors)

- *Biological Sciences* - Major Field Achievement Test (each year), Medical College Admissions Test (each year)
- *Center for Peaceful Change* - Field Study (final semester of junior year or senior year), Seminar in Methodology (senior year), Field Study Log/Final Paper (final semester of junior year or senior year)
- *Chemistry* - Major Field Achievement Test (each year), Department Survey of Seniors (each year to graduating students)

- *Classical Studies* - Master of Arts Comprehensive Examination (semester of graduation), Bachelor of Arts Senior Examination (semester of graduation), Undergraduate Major Portfolio (every semester), Alumni Questionnaire (each year)
- *Communication Studies* - Alumni Survey (every third year's students), National Communication Competence Assessment (graduating seniors), Exit Interview (all seniors)

- *Criminal Justice Studies* - Alumni Survey (Spring 1993 covering graduates from 1983-1992), External Review of Internship Students (each semester), Comprehensive Examination (under development)
- *Economics* - Major Field Achievement Test (each year), Survey of Graduating Seniors (each year), Intensive Writing Course (each semester to enrolled students), Alumni Tracking Program (under development), Note: Department of Economics uses the Major Field Achievement Test in Economics rather than the General Business Test because faculty determined that the General test did not cover enough ECON material.
- *Educational Psychology and Leadership Studies* - Annual Survey of Graduates (each year)
- *English: Non-Major Courses* - Developmental English Exit Examination (each semester), Writing Portfolio (each semester)
- *English: Undergraduate Majors* - Major Field Achievement Test (each year), Law School Admissions Test (each year), Graduate Record Examination (each year)
- English: Graduate Programs - MA and PhD Placement Survey (each year)

- *Family and Consumer Studies* - Alumni Survey (each Fall semester), Survey of Employers and Practicum Supervisors (each Fall semester), NCIDQ examination for Interior Designers (after graduation), ADA examination for Dietitians (after graduation and dietetic internships), License Examination for State and national Nursing Home Administrators (after graduation), License Examination for State and National Nursing Home Administration for undergraduate Gerontology majors (after graduation)
- *Fashion Design and Merchandising* - Alumni Survey (one time), Student Opinion Survey (one time), Sophomore Design Review (each year), Junior Portfolio Review (each year), Senior Portfolio Review (each year), Senior Student Fashion Show (each spring), Senior Seminar (Capstone course; each semester), Student Design Competitions (six to ten per year)
- *Finance* - Major Field Achievement Test (each year to Business majors), Graduating Senior Survey (each year)
- *Geography* - Senior Survey (each semester for graduating seniors), Alumni Survey (every two years to maintain continuous follow-up)

- *Geology* - Alumni Survey (each year), Employer Survey (at graduate for undergraduates and graduate students), Major Field Achievement Test (each year), Internships Evaluated by External Evaluators (all majors)
- *Germanic and Slavic Languages and Literatures* - Oral Proficiency Assessment (end of basic language sequence and seniors), Goethe Institute Language Proficiency Test (each year to German senior majors), Modern Languages Association Cooperative Language Test (at end of basic sequence or when senior), General Content Exam (under development, beginning Spring 1994), Oral Proficiency Assessment (entering graduate students), Final Translation Project and Oral Examination (Linguistics Masters students), Comprehensive Examinations (literature masters students), Goethe Institute Intermediate Language Proficiency Test (graduate masters students)
- *History* - Major Field Achievement Test (each year to graduating seniors), Departmental Student Satisfaction Survey (each year)
- *School of Music* - Major Field Achievement Test (graduating seniors), National Teacher Examination (Music Education Majors Only once each year to graduating seniors), Portfolio (composition majors at end of sophomore year, must pass to continue), Alumni Survey (under development), Senior Recital (performance majors), Functional Piano Proficiency (diagnostic examination, must be passed by all Music Education majors), Applied Music Juries (each semester, all students)

- *Journalism and Mass Communication* - Accreditation Review by the Accrediting Council for Education in Journalism and Mass Communication (every five years), Required Internships with External Review by Professional Supervisors (all majors), Exit Survey of Graduates (each semester), Diagnostic Writing Skills Test (all students prior to admission to professional courses), Advising Survey (once
- *Marketing* - Major Field Achievement Test (each semester to graduating seniors), Graduating Senior Survey (each year)
- *Mathematical and Computer Science* - Major Field Achievement Test (all graduating seniors in mathematics and computer science), Graduating Senior Academic Effectiveness Survey (each year), Alumni Survey (one year after graduation, comprehensive survey every five to eight years, undergraduate, masters, doctoral students), Career Follow-Up Survey (all masters and doctoral graduates), Publication Record of Theses and Dissertations (continuous)

- *Pan African Studies* - Department Outcomes Assessment Survey (given to new juniors and graduating seniors), Senior Colloquium Capstone course (seniors)
- *Philosophy* - Performance scores on Verbal and Analytic Skills test, using a GRE-like test (under development, will be given every semester), Writing Quality Score, using Grammatik-specified measures (under development, will be given every semester), Assessment of Student Portfolio consisting of philosophy papers submitted every semester (given every semester), Four Writing-Intensive Courses (each student)
- *Physics* - Major Field Achievement Test (each year to graduating seniors), Exit Survey (each graduating senior), Alumni Survey (after one year and at five-year intervals), Post-graduate placement (MS and PhD students), Publications and Presentations (MS and PhD students)
- *Political Science* - Major Field Achievement Test (each year to graduating seniors), Capstone course for majors (beginning in 1993-94), Law School Admission Test (offered four times a year for prospective law students)
- *Psychology* - Major Field Achievement Test (each year), Undergraduate Alumni Survey (each year), GRE scores (used when sample is sufficiently large to be reported), Placement Survey of PhD students in clinical and experimental psychology (each year), Performance of Clinical PhD students on state licensure examinations (from Placement Survey), Professional Scholarly Activity Prior to Final Orals (each year), Number of Clinical PhD students obtaining APA-approved internships (each year)

- *Romance Languages and Literatures* - American Council on the Teaching of Foreign Languages Oral Proficiency Interview (graduating seniors), Written Comprehensive Examination (graduating seniors), Comprehensive Examination or Thesis and Modified Comprehensive Examination (Masters literature students), Translation Thesis (Masters translation students), Teaching Demonstration on Video with Written Critique and Written and Oral Proficiency Tests (Masters pedagogy students), Oral Proficiency Interviews (each semester to students in fifth semester of language study), Alumni Questionnaire (under discussion)
- *Sociology* - Major Field Achievement Test (each year since 1990), GRE Specialty in Sociology (each year to participating seniors), Master of Arts Thesis Examination, Exit Survey (masters and doctoral students), Alumni survey (masters and doctoral students), Departmental Qualifying Tests in Theory and Statistics (doctoral students), Departmental Methods Test (doctoral level entrants), dissertation examination (doctoral students)
- *Speech Pathology and Audiology* - Departmental Questionnaire (graduating seniors), GPA in major (all students by semester), National Examinations in Speech-Language Pathology and Audiology (masters and doctoral students), Student Practicum Evaluation by Clinical Supervisor (masters and doctoral students), American Speech Language Hearing Association accreditation-compliance annual report (each year for the program)

- *Teacher Development and Curriculum Studies* - Pre-Professional Skills Test (prior to admission to advanced study), Written Assessment of Communication Skills and Teaching Characteristics (prior to admission to advanced study), National Teacher Examination (all seniors and graduate students obtaining certification), Structured Observations in Field Settings (each semester), Survey of Program Satisfaction (during Student Teaching), Structured Observations at Field Sites by External Observers

(each semester), Portfolio (end of study in some programs), Employment Survey (each year undergraduate and graduate students), Satisfaction with Program and Professional Preparation (one year and three years after graduation)

- *Technology: Aerospace Flight Technology* - Each Aerospace Flight Technology major is required to have five FAA licenses as part of the degree program. Sequences of courses are taken which include both flight and ground school for each license. (At or near the end of each sequence of courses, the FAA examination in both the written and flight modes are taken. Students cannot complete the degree requirements without these licenses.), Alumni Survey (every fifth year), Employer Survey (every fifth year), Exit Interview (under consideration)
- *Technology: Technology Education* - National Teacher Examination (seniors, each year), Alumni Survey (every fifth year), Employer Survey (every fifth year), Exit Interview (under consideration)
- *Theatre* - Public presentation within School (once each semester), Juried Evaluation (BFA seniors, each year), participation in campus theatre (six public productions per year, students participate on a competitive basis), Participation in American College Theatre Festival (national competition, annual), Alumni Survey (every fifth year)

ATTACHMENT 4. Cautions/Advice for and Illustrative Examples of Selected Assessment Methods

Case Studies
- The study of a single case should not be haphazard and unstructured
- We should always be cautious about making generalizations. The broader the sample of cases, the more confidence in the findings
- Case studies often involve observing or studying only one subject at a given time, so results may not generalize beyond the specific observed situation
- Keep your objectives as you select and analyze the case.

Content Analysis
- Summaries based on concrete materials and coding schemes are more reliable
- No coding scheme should be used unless it has been carefully pre-tested
- It is essential that the coding scheme distinguish between units of analysis (e.g., students) and units of observation (e.g., paragraphs with documents)
- Summarize the extent of agreement or similarity among respondents numerically (e.g., 40% of the students reported problems locating relevant articles in the library)
- When providing counts, report the base from which the counting is done (e.g., 3.2 spelling errors per page)
- Compare the results of two or more independent analysts examining the same documents to check the reliability of the coding scheme

Focus groups
- Clearly identify the focus groups goals through discussions with program stakeholders
- Carefully select those you invite and encourage their participation -- recruiting the right participants is essential
- Anticipate what kind of information you want in the final report and be sure to include questions and participants that will allow you to obtain that information
- Do a pilot focus group to determine the effectiveness of your approach
- Record sessions on audio/video and type transcripts for subsequent analysis
- Conduct more than one group to test for the consistency of results across groups
- Use more than one data analyst as a check on the reliability of the coding process

Suggestions for Focus Groups
1. What does the program do well? What are its greatest strengths in the eyes of students? How could the program made stronger?
2. What aspects of the curriculum do students consider most important for their careers? For grad school? For life?
3. What are the most serious obstacles to student success in your program? How might these obstacles be removed or minimized?
4. What is the experience of students when they first enter the program? What could make the transition more effective?
5. What types of students are best/worst served by the program? How can we serve them better?
6. Do students have access to needed program information in timely fashion?
7. What skills and knowledge do employers most want in your graduates? How well does the program provide opportunities to learn the desired skills and knowledge?
8. What are things about your work situation that make it difficult to perform as you would like?
9. If you could change one thing about your work situation that would help you do a better job, what would it be?

Interviews
- Shorter interviews require less time and are more likely to gain student participation

- An impressive letter from the campus president may encourage participation
- Students who agree to be interviewed may need to be reminded by phone
- Student schedules and pressures must be considered in developing assessment plan. Do not schedule interviews during exam periods or vacations
- Interviewers need training in interview techniques
- Every question should serve a purpose
- Remember that the process is to evaluate the program, not individuals
- Avoid setting up situations with strong demand characteristics that may distort the types of responses you get from interviewee (e.g., professors should not interview students who are taking or will take classes from them)
- Avoid judgmental or evaluative statements which are likely to inhibit the interviewee
- Replace "Why?" questions with "Tell me…" or "How did it happen that…" questions
- Do not give false reassurance
- Remember to respect confidentiality and the right to decline participation

Examples:
1. Truman State University. Freshmen and juniors were each interviewed by two co-interviewers using a 20 minute structured interview, with questions like " What is a challenging course?" Students were paid with lunch.
2. Kansas State U. Seniors were interviewed in groups (each was paid $25), and three faculty interviewed individual students for 45-50 minutes.
3. U. of Kansas. Students were interviewed for 45-50 minutes to assess their general education program
4. Ball State U. A student panel was interviewed several times per year (Reflection and Assessment Panel). Students were paid up to $350 per year.
5. Portland State U. Student, faculty, and community interviews were used to assess the impact of service learning.
6. Alverno College. Student self-assess with videotaped interviews at several points during college.

Locally-Developed Exams: Essay Questions
- Be sure that questions are clearly phrased so student writing will be focused on your objective
- Consider pilot testing your essay questions on relevant students and faculty
- Examine the reliability and validity of your scores
- Consider using Blooms' taxonomy (see Attachment 1). Do your essay questions address relevant levels?

Types of Essay Questions.
1. Compare and contrast X and Y in regard to given qualities
2. Present arguments for and against a given issue
3. Illustrate how a principle explains facts
4. Illustrate cause and effect
5. Describe an application of a rule or principle
6. Evaluate the adequacy, relevance, or implication of an arrangement, or materials, etc.
7. Form new inferences from data
8. Organize the parts of situation, event, or mechanism and show how they interrelate into a whole
9. Sort out the relevant parts as distinct entities form a total situation event or mechanism

Key words in essay questions are *summarize, evaluate, contrast, explain, describe, define, compare, discuss, criticize, justify, trace, interpret, prove,* and *illustrate*.

A formula for writing essay questions generally involves three parts: *a role, an audience, and a task*. For example, "As a certified financial planner [the role] you are asked to explain a jury [the audience] how to estimate a thirty-year –old carpenter's loss of lifetime earnings after an accident [the task]. Be sure to take into account savings, investments, inflation, and post-retirement earnings. The carpenter is no longer able to lift anything over ten pounds and is hoping to find work in retail sales."
Locally-Developed Exams: Objective Questions

Types of Items
- Multiple-Choice Questions. Multiple-choice questions can measure many objectives in a short period of time and are better than true-false items because the chance of correct guess is less than 50%.
  Example:
  > Jacob's literature review clearly supports the effectiveness of a new sleeping pill, but he did not get significant results in his study of 3 research subjects. What is the most reasonable conclusion?
  > a.   the drug is not effective
  > b.   Jacob probably made a Type 1 error
  > c.   Jacob's study lacked sufficient power to reveal the drug's effect
  > d.   Jacob's study probably had restriction of range

- Matching items. This is a group of multiple-choice questions with the same set of possible answers.
  Example:
  > Mark the letter of the word best described by the following
  > __1. an indicator of central tendency   A. variance
  > __2. the square of the standard deviation   B. t-test
  > __3. used to compare two means   C. Spearman
  > __4. a type of correlation   D. mean
  > __5. a two-parameter correlation   E. normal

- Completion items. Completion items must be developed carefully do a unique answer is correct. These items require recall, rather than recognition.
  Examples:
  > 1.   The _____is the square of the standard deviation
  > 2.   The two major issues of inferential statistics are _____and _____.

Matrices
- Discussion of the matrices should involve all relevant faculty in a collegial, consensual conversation
- Important learning goals generally should be introduced early in the major, practiced repeatedly, and reinforced by later integration into advanced course
- Matrices can just contain checkmarks, or they can contain more detailed indicators of level,
  Examples:
  Course x Program Goals Matrix

Course	Goal 1	Goal 2	Goal 3	Goal 4	Goal 5
100	I				I
120		I			P
200	P		P		P
204					P
300	P		P		
329					P
400					R
480					
490	R		R		R

I-introduced; P-practiced; R-reinforced

This program fails to practice and reinforce Goal 2, so graduating seniors may not have mastered it. Goal 3 was never formally introduced. Goal 4 was not included in the curriculum.

Course Assignment x Intended Program Outcome

Assignment	Outcome 1	Outcome 2	Outcome 3	Outcome 4
1	B			
2	I		B	
3	I			I
4	A			I

B-basic, I-intermediate, A-advanced expectation for this outcome

Observations

- For program assessment, purposive sampling may be the most efficient
- Sometimes note taking can be simplified by preparing a standardized recording form in advance
- If the observer is a faculty member in the program being assessed, it is unlikely that she or he can be either complete observer or an observer as participant, she her/his presence may affect the process being observed
- Unstructured interviews are usually more appropriate to participant observation studies.

Phone Surveys / Interviews
- The survey generally should not take more than 20 minutes to complete
- A very structured interview should be used, a questions should be asked at a reasonably quick pace so respondents do not get bored
- The issue of interest should be well defined and articulated
- Although closed-ended questions are generally more desirable, it is possible to include a very limited number of open-ended items
- Probability sampling designs are essential for obtaining reliable and valid data that generalize to populations of interest.

Portfolios
- Anticipate what you want the portfolio to tell you about the program and be sure to structure the portfolio assignment to provide that information
- Do not expect reliable and valid results if students do not understand the process or the rationale. Provide guidance through handouts and advising, and create a culture of understanding by embedding portfolio awareness throughout the curriculum
- Student self-evaluation and program evaluation should be an integral part of the portfolio
- If you want to assess student progress, ask for early and late examples of products so that change can be observed
- Reduce costs by setting priorities (e.g., assess a limited number of outcomes each year and/or assess a sample of students rather than all students in the program)
- Increase faculty motivation to participate by recognizing the portfolio process in the workload or by other incentives
- Increase student motivation by providing credit or pay for participation, or by encouraging students to recognize the educational and career value of the process for them
- Develop a holistic scoring rubric with a written scoring protocol and examples of different levels of performance. This allows faculty to assess portfolios more efficiently and reliably, and it provides a way of communicating to others what each level of performance means. Share the rubric with students.

Reflective Essays
Examples
- Before you came to NSU, what did you think college would be like? [what expectations do students bring to NSU?]
- Write about the ways in which your NSU experience has changed or has not changed your thinking about college. [a quick look at impact]
- What have you read, observed, heard, or done in the past semester that caused you to recognize and examine your assumptions about people different from yourself? [diversity]
- Share what you liked best about your classes last semester and what you liked least. [pedagogy feedback]
- Describe a course assignment that asked you to identify and work on a question, issue, or problem. [critical thinking assignments]
- Describe as assignment that asked you to collaborate with other students on a project. To what extent did working with others help or hinder your learning? [collaborative learning impact]
- What did you expect to gain from being a ***major and were expectations met? [match between intended learning outcomes identified by faculty and students]

Transcript Analysis

Patterns

- Curriculum sequencing. Did students take courses in the expected order? Did deviations from the expected pattern result in lower grades? Should formal prerequisite sequences or better advising be instituted?
- General Education class order. When were GenEd goals taken? Did delaying one of the basic goals affect achievement in other course, persistence, or graduation?
- Possible prerequisite sequences. Did students who took English 200 before taking Psychology 100 do better? Should the department consider making English 200 a prerequisite?
- Transfer students. How do they differ from native students in upper-division coursework in the major? Should a transition course, experience, or competency exam be created?
- Drop-outs vs. Stop-outs vs. Graduates. What course patterns, if any, distinguish between these three groups?
- Standardized test scores. Do SAT or ACT scores relate to success in the major?
- Graduate program success. Are successful graduate students different from unsuccessful graduate students in their academic histories?
- Type of student. How many of your majors are full-time vs. part-time? What proportion are transfer students? What proportion has not completed relevant GenEd requirements?
- Needed courses. How many majors are cued up to take Course X? How many sections must be scheduled for this course next year?
- Outcomes data. How do transcript records relate to the performance on core competencies assessments?

Example

Xs show when a student declared the major and graduated, and the overall pattern should be from the upper left to the lower right if courses are correctly sequenced. Students with deviations from this pattern can be identified at a glance and comparisons can be made to see if deviations resulted in reduced performance. For example, the student below took 300 before taking 201 and had an unusually low grade. If this pattern were consistent across students, the department might consider adding 201 as a prerequisite for 300.

Semester/ Course	1	2	3	4	5	6	7	8	9
declared				X					
100	A								
200		B							
201				A					
300			D						
312						A			
339					A				
400					B				
425							A		
490							A		
graduated									X