Measurement, Assessment, and Evaluation in Education

Dr. Bob Kizlik

Updated January 4, 2012

Throughout my years of teaching undergraduate courses, and to some extent, graduate courses, I was continuously reminded each semester that many of my students who had taken the requisite course in "educational tests and measurements" or a course with a similar title as part of their professional preparation, often had confusing ideas about fundamental differences in terms such as measurement, assessment and evaluation as they are used in education. When I asked the question, "what is the difference between assessment and evaluation," I usually got a lot of blank stares. Yet, it seems that understanding the differences between measurement, assessment, and evaluation is fundamental to the knowledge base of professional teachers and effective teaching. Such understanding is also, or at the very least should be a core component of the curricula implemented in universities and colleges required in the education of future teachers.

In many places on the ADPRIMA website the phrase, "Anything not understood in more than one way is not understood at all" appears after some explanation or body of information. That phrase is, in my opinion, a fundamental idea of what should be a cornerstone of all teacher education. Students often struggle with describing or explaining what it means to "understand" something that they say they understand. I believe in courses in educational tests and measurements, that "understanding" has often been inferred from responses on multiple-choice tests or solving statistical problems. A semester later, when questioned about very fundamental ideas in statistics, measurement, assessment and evaluation, the students I had seemingly forgot most, if not all of what they "learned."

Measurement, assessment, and evaluation mean very different things, and yet most of my students were unable to adequately explain the differences. So, in keeping with the ADPRIMA approach to explaining things in as straightforward and meaningful a way as possible, here are what I think are useful descriptions of these three fundamental terms. These are personal opinions, but they have worked for me for many years. They have operational utility, and therefore may also be useful for your purposes.

Measurement refers to the process by which the attributes or dimensions of some physical object are determined. One exception seems to be in the use of the word measure in determining the IQ of a person. The phrase, "this test measures IQ" is commonly used. Measuring such things as attitudes or preferences also applies. However, when we measure, we generally use some standard instrument to determine how big, tall, heavy, voluminous, hot, cold, fast, or straight something actually is. Standard instruments refer to instruments such as rulers, scales, thermometers, pressure gauges, etc. We measure to obtain information about what is. Such information may or may not be useful, depending on the accuracy of the instruments we use, and our skill at using them. There are few such instruments in the social sciences that approach the validity and reliability of say a 12" ruler. We measure how big a classroom is in terms of square feet, we measure the temperature of the room by using a thermometer, and we use Ohm meters to determine the voltage, amperage, and resistance in a circuit. In all of these examples, we are not assessing anything; we are simply collecting information relative to some established rule or standard. Assessment is therefore quite different from measurement, and has uses that suggest very different purposes. When used in a learning objective, the definition provided on the ADPRIMA for the behavioral verb measure is: To apply a standard scale or measuring device to an object, series of objects, events, or conditions, according to practices accepted by those who are skilled in the use of the device or scale.

Assessment is a process by which information is obtained relative to some known objective or goal. Assessment is a broad term that includes testing. A test is a special form of assessment. Tests are assessments made under contrived circumstances especially so that they may be administered. In

other words, all tests are assessments, but not all assessments are tests. We test at the end of a lesson or unit. We assess progress at the end of a school year through testing, and we assess verbal and quantitative skills through such instruments as the SAT and GRE. Whether implicit or explicit, assessment is most usefully connected to some goal or objective for which the assessment is designed. A test or assessment yields information relative to an objective or goal. In that sense, we test or assess to determine whether or not an objective or goal has been obtained. Assessment of skill attainment is rather straightforward. Either the skill exists at some acceptable level or it doesn't. Skills are readily demonstrable. Assessment of understanding is much more difficult and complex. Skills can be practiced; understandings cannot. We can assess a person's knowledge in a variety of ways, but there is always a leap, an inference that we make about what a person does in relation to what it signifies about what he knows. In the section on this site on behavioral verbs, to assess means To stipulate the conditions by which the behavior specified in an objective may be ascertained. Such stipulations are usually in the form of written descriptions.

Evaluation is perhaps the most complex and least understood of the terms. Inherent in the idea of evaluation is "value." When we evaluate, what we are doing is engaging in some process that is designed to provide information that will help us make a judgment about a given situation. Generally, any evaluation process requires information about the situation in question. A situation is an umbrella term that takes into account such ideas as objectives, goals, standards, procedures, and so on. When we evaluate, we are saying that the process will yield information regarding the worthiness, appropriateness, goodness, validity, legality, etc., of something for which a reliable measurement or assessment has been made. For example, I often ask my students if they wanted to determine the temperature of the classroom they would need to get a thermometer and take several readings at different spots, and perhaps average the readings. That is simple measuring. The average temperature tells us nothing about whether or not it is appropriate for learning. In order to do that, students would have to be polled in some reliable and valid way. That polling process is what evaluation is all about. A classroom average temperature of 75 degrees is simply information. It is the context of the temperature for a particular purpose that provides the criteria for evaluation. A temperature of 75 degrees may not be very good for some students, while for others, it is ideal for learning. We evaluate every day. Teachers, in particular, are constantly evaluating students, and such evaluations are usually done in the context of comparisons between what was intended (learning, progress, behavior) and what was obtained. When used in a learning objective, the definition provided on the ADPRIMA site for the behavioral verb evaluate is: To classify objects, situations, people, conditions, etc., according to defined criteria of quality. Indication of quality must be given in the defined criteria of each class category. Evaluation differs from general classification only in this respect.

To sum up, we measure distance, we assess learning, and we evaluate results in terms of some set of criteria. These three terms are certainly connected, but it is useful to think of them as separate but connected ideas and processes.

Here is a great link that offer different ideas about these three terms, with well-written explanations. Unfortunately, most information on the Internet concerning this topic amounts to little more than advertisements for services.

The definitions for each are:

Test: A method to determine a student's ability to complete certain tasks or demonstrate mastery of a skill or knowledge of content. Some types would be multiple choice tests, or a weekly spelling test. While it is commonly used interchangeably with assessment, or even evaluation, it can be distinguished by the fact that a test is one form of an assessment.

Assessment: The process of gathering information to monitor progress and make educational decisions if necessary. As noted in my definition of test, an assessment may include a test, but also includes methods such as observations, interviews, behavior monitoring, etc.

Evaluation: Procedures used to determine whether the subject (i.e. student) meets a preset criteria, such as qualifying for special education services. This uses assessment (remember that an assessment may be a test) to make a determination of qualification in accordance with a predetermined criteria.

Measurement, beyond its general definition, refers to the set of procedures and the principles for how to use the procedures in educational tests and assessments. Some of the basic principles of measurement in educational evaluations would be raw scores, percentile ranks, derived scores, standard scores, etc.

Reference:

Assessing Learners with Special Needs: 6TH ED. By Terry Overton

Assessment, Evaluation, Testing and Grading

Craig L. Scanlan, EdD, RRT

What is Assessment?

To many teachers (and students), "assessment" simply means giving students tests and assigning them grades. This conception of assessment is not only limited, but also limiting (see section below on Assessment versus grading). It fails to take into account both the utility of assessment and its importance in the teaching/learning process.

In the most general sense, *assessment* is the process of making a judgment or measurement of worth of an entity (e.g., person, process, or program). *Educational assessment* involves gathering and evaluating data evolving from planned learning activities or programs. This form of assessment is often referred to as *evaluation* (see section below on Assessment versus Evaluation). *Learner assessment* represents a particular type of educational assessment normally conducted by teachers and designed to serve several related purpose (Brissenden and Slater, n.d.). These purposed include:

- motivating and directing learning
- providing feedback to student on their performance
- providing feedback on instruction and/or the curriculum
- ensuring standards of progression are met

Learner assessment is best conceived as a form of two-way communication in which feedback on the educational process or product is provided to its key stakeholders (McAlpine, 2002). Specifically, learner assessment involves communication to **teachers** (feedback on teaching); **students** (feedback on learning); **curriculum designers** (feedback on curriculum) and **administrators** (feedback on use of resources).

For teachers and curriculum/course designers, carefully constructed learner assessment techniques can help determining whether or not the stated goals are being achieved. According to Brissenden and Slater (n.d.), classroom assessment can help teachers answer the following specific questions:

- To what extent are my students achieving the stated goals?
- How should I allocate class time for the current topic?
- Can I teach this topic in a more efficient or effective way?
- What parts of this course/unit are my students finding most valuable?

- How will I change this course/unit the next time I teach it?
- Which grades do I assign my students?

For students, learner assessment answers a different set of questions (Brissenden and Slater, n.d.):

- Do I know what my instructor thinks is most important?
- Am I mastering the course content?
- How can I improve the way I study in this course?
- What grade am I earning in this course?

Why Assessment is Important

First and foremost, **assessment is important because it drives students learning** (Brissenden and Slater, n.d.). Whether we like it or not, most students tend to focus their energies on the best or most expeditious way to pass their 'tests.' Based on this knowledge, we can use our assessment strategies to manipulate the kinds of learning that takes place. For example, assessment strategies that focus predominantly on recall of knowledge will likely promote superficial learning. On the other hand, if we choose assessment strategies that demand critical thinking or creative problem-solving, we are likely to realize a higher level of student performance or achievement. In addition, good assessment can help students become more effective self-directed learners (Angelo and Cross, 1993).

As indicated above, motivating and directing learning is only one purpose of assessment. Well-designed assessment strategies also play a critical role in educational decision-making and are a vital component of ongoing quality improvement processes at the lesson, course and/or curriculum level.

Types and Approaches to Assessment

Numerous terms are used to describe different types and approaches to learner assessment. Although somewhat arbitrary, it is useful to these various terms as representing dichotomous poles (McAlpine, 2002).

Formative	<------------------------------->	Summative
Informal	<------------------------------->	Formal
Continuous	<------------------------------->	Final
Process	<------------------------------->	Product
Divergent	<------------------------------->	Convergent

Formative vs. Summative Assessment

Formative assessment is designed to assist the learning process by providing feedback to the learner, which can be used to identify strengths and weakness and hence improve future performance. Formative assessment is most appropriate where the results are to be used internally by those involved in the learning process (students, teachers, curriculum developers).

Summative assessment is used primarily to make decisions for grading or determine readiness for progression. Typically summative assessment occurs at the end of an educational activity and is designed to judge the learner's overall performance. In addition to providing the basis for grade assignment, summative assessment is used to communicate students' abilities to external stakeholders, e.g., administrators and employers.

Informal vs. Formal Assessment

With informal assessment, the judgments are integrated with other tasks, e.g., lecturer feedback on the answer to a question or preceptor feedback provided while performing a bedside procedure. Informal assessment is most often used to provide formative feedback. As such, it tends to be less threatening and thus less stressful to the student. However, informal feedback is prone to high subjectivity or bias.

Formal assessment occurs when students are aware that the task that they are doing is for assessment purposes, e.g., a written examination or OSCE. Most formal assessments also are summative in nature and thus tend to have greater motivation impact and are associated with increased stress. Given their role in decision-making, formal assessments should be held to higher standards of reliability and validity than informal assessments.

Continuous vs. Final Assessment

Continuous assessment occurs throughout a learning experience (intermittent is probably a more realistic term). Continuous assessment is most appropriate when student and/or instructor knowledge of progress or achievement is needed to determine the subsequent progression or sequence of activities. Continuous assessment provides both students and teachers with the information needed to improve teaching and learning *in process*. Obviously, continuous assessment involves increased effort for both teacher and student.

Final (or terminal) assessment is that which takes place only at the end of a learning activity. It is most appropriate when learning can only be assessed as a complete whole rather than as constituent parts. Typically, final assessment is used for summative decision-making. Obviously, due to its timing, final assessment cannot be used for formative purposes.

Process vs. Product Assessment

Process assessment focuses on the steps or procedures underlying a particular ability or task, i.e., the cognitive steps in performing a mathematical operation or the procedure involved in analyzing a blood sample. Because it provides more detailed information, process assessment is most useful when a student is learning a new skill and for providing formative feedback to assist in improving performance.

Product assessment focuses on evaluating the result or outcome of a process. Using the above examples, we would focus on the answer to the math computation or the accuracy of the blood test results. Product assessment is most appropriate for documenting proficiency or competency in a given skill, i.e., for summative purposes. In general, product assessments are easier to create than product assessments, requiring only a specification of the attributes of the final product.

Divergent vs. Convergent Assessment

Divergent assessments are those for which a range of answers or solutions might be considered correct. Examples include essay tests, and solutions to the typical types of indeterminate problems posed in PBL. Divergent assessments tend to be more authentic and most appropriate in evaluating higher cognitive skills. However, these types of assessment are often time consuming to evaluate and the resulting judgments often exhibit poor reliability.

A convergent assessment has only one correct response (per item). Objective test items are the best example and demonstrate the value of this approach in assessing knowledge. Obviously, convergent assessments are easier to evaluate or score than divergent assessments. Unfortunately, this "ease of use" often leads to their widespread application of this approach even when contrary to good assessment practices. Specifically, the familiarity and ease with which convergent assessment tools can be applied leads to two common evaluation fallacies: the Fallacy of False Quantification (the tendency to focus on what's easiest to measure) and the Law of the Instrument Fallacy (molding the evaluation problem to fit the tool).

Assessment versus Evaluation

Depending on the authority or dictionary consulted, assessment and evaluation may be treated as synonyms or as distinctly different concepts. As noted above, if a distinction exists, it probably involves what is being measured and why and how the measurements are made. In terms of what, it is often said that we assess students and we evaluate instruction. This distinction derives from the use of *evaluation research* methods to make judgments about the worth of educational activities. Moreover, it emphasizes an individual focus of assessment, i.e., using information to help identify a learner's needs and document his or her progress toward meeting goals.

In terms of why and how the measurements are made, the following table (Apple & Krumsieg, 1998) compares and contrasts assessment and evaluation on several important dimension, some of which were previously defined.

Dimension	Assessment	Evaluation
Timing	Formative	Summative
Focus of Measurement	Process-Oriented	Product-Oriented
Relationship Between Administrator and Recipient	Reflective	Prescriptive
Findings and Uses	Diagnostic	Judgmental
Modifiability of Criteria, Measures	Flexible	Fixed
Standards of Measurement	Absolute (Individual)	Comparative
Relation Between Objects of A/E	Cooperative	Competitive

From: Apple, D.K. & Krumsieg. K. (1998). *Process education teaching institute handbook*. Pacific Crest

The bottom line? Given the different meaning ascribed to these terms by some educators, it is probably best that whenever you use these terms, you make *your* definitions clear.

Assessment versus Grading

Based on the above discussion, grading grading could be considered a component of assessment, i.e., a formal, summative, final and product-oriented judgment of overall quality of worth of a student's performance or achievement in a particular educational activity, e.g., a course. Generally, grading also employs a comparative standard of measurement and sets up a competitive relationship between those receiving the grades. Most proponents of assessment, however, would argue that grading and assessment are two different things, or at least opposite pole on the evaluation spectrum. For them, assessment measures student growth and progress on an individual basis, emphasizing informal, formative, process-oriented reflective feedback and communication between student and teacher. Ultimately, which conception you supports probably depends more on your teaching philosophy than anything else.

Bibliography

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

Apple, D.K, & Krumsieg. K. (1998). *Process education teaching institute handbook*. Corvalis, OR: Pacific Crest Software.

Brissenden, G., & Slater, T. Assessment primer. In College Level One (CL-1) Team. *Field-tested learning assessment guide*. Available at http://www.flaguide.org.

Linn, R. L. (1995). *Measurement and assessment in teaching* (7th ed.). Englewood Cliffs, NJ: Merrill.

McAlpine, M. (2002). *Principles of assessment*. Glasgow: University of Glasgow, Robert Clark Center for Technological Education. Available at: http://www.caacentre.ac.uk/dldocs/Bluepaper1.pdf

Wiggins, G. P. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass.

Wass, V., Van der Vleuten, J., & Shatzer, R.J. (2001). Assessment of clinical competence. *The Lancet, 357*, 945-949.

http://specialed.about.com/od/assessment/a/AandE.htm

What's the Difference between Assessment, Evaluation and Final Marks or Report Card Grades?

The overall goal of assessment is to *improve student learning.* Assessment provides students, parents/guardians, and teachers with valid information concerning student progress and their attainment of the exptected curriculum/IEP. Assessment should always be viewed as information to improve student achievement. Assessments are based on the levels of achievement and standards developed for those curricular goals appropriate for the grade or those listed in the IEP. Assessment and evaluation measure whether or not learning and/or learning objectives are being met. One could look at assessment and evaluation as the journey (assessment) versus the snapshot (evaluation). Assessment requires the gathering of evidence of student performance over a period of time to measure learning and understanding. Evidence of learning could take the form of dialogue, journals, written work, portfolios, tests along with many other learning tasks. Evaluation on the other hand occurs when a mark is assigned after the completion of a task, test, quiz, lesson or learning activity. A mark on a spelling test will determine if the child can spell the given words and would be seen as an evaluation. Assessment would be a review of journal entries, written work, presentation, research papers, essays, story writing, tests, exams etc. and will demonstrate a sense of more permanent learning and clearer picture of a student's ability. Although a child may receive high marks in spelling test, if he/she can't apply correct spelling in every day work, the high spelling test marks (evaluations) matter little.

Effective teachers will use both assessment and evaluation techniques regularly and on a daily basis to improve student learning and to guide instruction.

Assessment, Evaluation, and Report Card Marks or Final Grades

Ongoing assessment and certain evaluations will make up final marks and/or report card grades. For instance, let's say I am ready to give a final/report card mark for language/English. Here is one example on how I would come up with that mark:

- 15% for notes done in class
- 10% for participation in group work
- 5% for homework completion
- 20% for a book report which was marked for content, conventions, grammar, process, understanding and spelling
- 20% which reflects the mean average of 3 quizzes given
- 20% for an oral presentation for which I was evaluating reasoning, oral communication and organization
- 10% Weekly spelling and grammar quizzes

The most important element of assessment and evaluation is to provide information for improved student performance. When one looks at a mark of 65% or a C in Language/English, it is important to know how the student can improve to receive a 75% or a B next time. Look at the evidence and work with the teacher to determine the areas of strength and the areas of weakness to improve overall learning

Alternatives in Assessment

by *Ali Bastanfar*

Islamic Azad University, Khoy, Iran

Introduction

The term *alternative assessment,* and particular testing practices associated with it, have recently come into vogue in language testing. The movement is directed at establishing qualitative, more democratic, and task-based methods of evaluation in testing a learner's language proficiency (Brown and Hudson 1998), (Aschbacher 1991), (Herman, Aschbacher, and Winters 1992), (Huerta-Macías 1995). It contrasts with traditional methods of testing by involving the learners in the evaluation process, and having the tendency to locate evaluation in a real-life context and, as result of these two features, being longitudinal. Thus, the insights emanating from these methods, alongside being used for decision-making about the future of learners, contribute to and furnish additional instructional purposes. As McNamara (2000) points out:

> *"This approach stresses the need for assessment to be integrated with the goals of the curriculum and to have a constructive relationship with teaching and learning".*

The procedures used within this paradigm include checklists, journals, logs, videotapes and audiotapes, self-evaluation, teacher observations, portfolios, conferences, diaries, self-assessments and peer-assessments (Brown and Hudson 1998). These procedures have been diversely called alternative or performance assessment as opposed to traditional assessment techniques such as multiple choice, cloze test, dictation, etc.

While the new movement promises more humanistic and rewarding methods of testing and thus has a lot to offer, most teachers are not quite familiar with the new concepts and practices within the emerging paradigm. To enlighten the views of interested teachers, it can be a good start to answer a basic question about the so-called alternative methods of testing which may have occupied their minds. This question is concerned with the relationship of these *other* methods with the traditional methods normally used within classrooms. Or to put the question another way, how can we place both traditional and alternative assessment methods in perspective to get a panoramic view of both in the pieced together jigsaw of language testing? To this purpose, it seems necessary to draw on the concepts of *testing*, *measurement* and *evaluation*.

Evaluation, Measurement and Testing

Bachman (1990), quoting Weiss (1972) defines evaluation as "the systematic gathering of information for the purpose of making decisions". Lynch (2001) adds the fact that this decision or judgment is to be about individuals. In this conceptualization, both authors agree that evaluation is the superordinate term in relation to both measurement and testing. Assessment is sometimes used interchangeably for evaluation. The systematic information can take many forms, but these forms are either quantitative or qualitative. This is what distinguishes measures from qualitative descriptions.

Measurement is thus concerned with quantification. Language proficiency, like many other constructs and characteristics of persons in social sciences, needs to be quantified before any judgments can be made about it. This process of quantifying is called operationalization in research by which we mean assigning numbers according to observable operations and explicit procedures or rules to measure a construct (Bachman 1990) (Ary et al. 1996)

The third component in this model is testing, which consists of the use of actual tests to elicit the desired behavior. Carroll (1968) defines a test as:

> "A psychological or educational test is a procedure designed to elicit certain behavior from which one can make inferences about certain characteristics of an individual".

Bachman (1990) observes that a test is one type of measurement instrument, and thus necessarily quantifies characteristics of individuals according to explicit procedures. Bachman (1990), then, concludes that there are other types of measurement than tests, and the difference is that a test is designed to obtain a specific sample of behavior.

For the purpose of schematic representation, the three concepts of evaluation, measurement and testing have traditionally been demonstrated in three concentric circles of varying sizes. This is what Lynch (2001) has followed in depicting the relationship among these concepts.
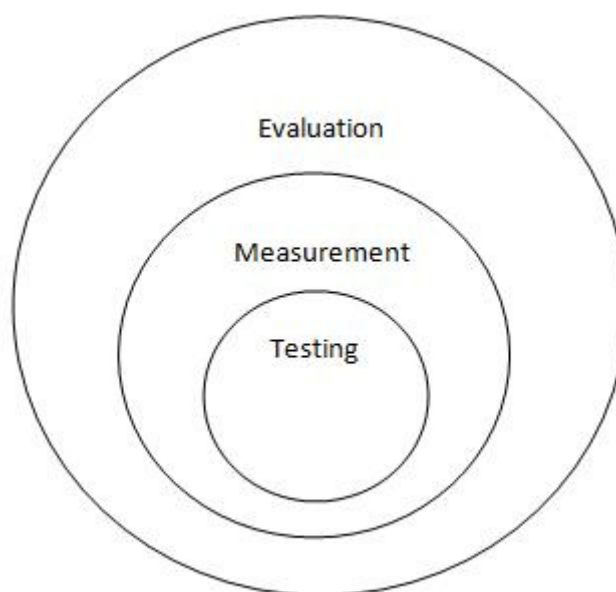
The purpose of this representation is to show the relationship between superordinate and subordinate concepts and the area of overlap between them. Thus, evaluation includes measurement when decisions are made on the basis of information from quantitative methods. And measurement includes testing when decision-making is done through the use of "a specific sample of behavior" (Bachman 1990). However, the process of decision-making is by no means restricted to the use of quantitative methods as the area not covered by measurement circle shows. Also, tests are not the only means to measure individuals' characteristics as there are other types of measurement than tests, for example, measuring an individual's language proficiency by living with him for a long time.

Bachman (1990) has represented the relationship in a somewhat different way. The goal has been to extend the model to include not only language testing but also language teaching, language learning and language research domains. Figure 2 depicts this extended view of the relationship among evaluation, measurement and testing. The areas numbered from 1 to 5 show the various forms of this relationship.
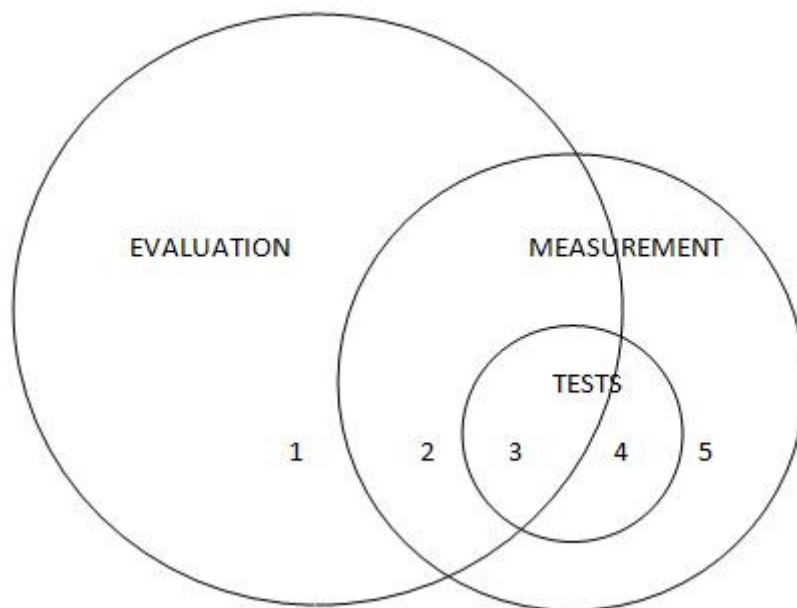

Figure 2- Assessment, measurement and testing adopted from Bachman (1990)

Area 1- Evaluation not involving either tests or measures; for example, the use of qualitative descriptions of student performance for diagnosing learning problems.

Area 2- A non-test measure for evaluation; for example, teacher ranking used for assigning grades.

Area 3- A test used for purposes of evaluation; for example, the use of an achievement test to determine student progress.

Area 4- Non-evaluative use of tests and measures for research purposes; for example, the use of a proficiency test as a criterion in second language acquisition research.

Area 5- Non-evaluative non-test; for example, assigning code numbers to subjects in second language research according to native language.

After reviewing the conceptualizations and schematic representations proposed by Bachman (1990) and Lynch (2001), an attempt will be made to more clearly locate alternative assessment methods in relation to traditional testing methods in order to help language teachers to make intelligent and insightful choices to assess their students. Some points are notable about the adapted model. First, despite Bachman's model, language research purposes are not dealt with in it. This is because language teachers' immediate needs do not concern the use of tests or assessment procedures for research purposes. Rather, they need to enhance their assessment choices to arrive at a sounder judgment about their students. Secondly, all assessment procedures either traditional or alternative furnish the function of decision-making and are all subordinated under the term evaluation. Thus, it would be much better to deal with them as *alternatives in assessment* (Brown and Hudson 1998) – available choices for the language teacher – rather than labeling some of them as normal and others as eccentric. Such a distinction makes the new developments inaccessible only because they are told to be so, hence our use of more descriptive terms instead of labels which evoke vague feelings. We have to notice the fact that all alternatives in assessment have to meet their respective requirements for reliability and validity to make teachers able to come to sound judgments (Lynch 2001).
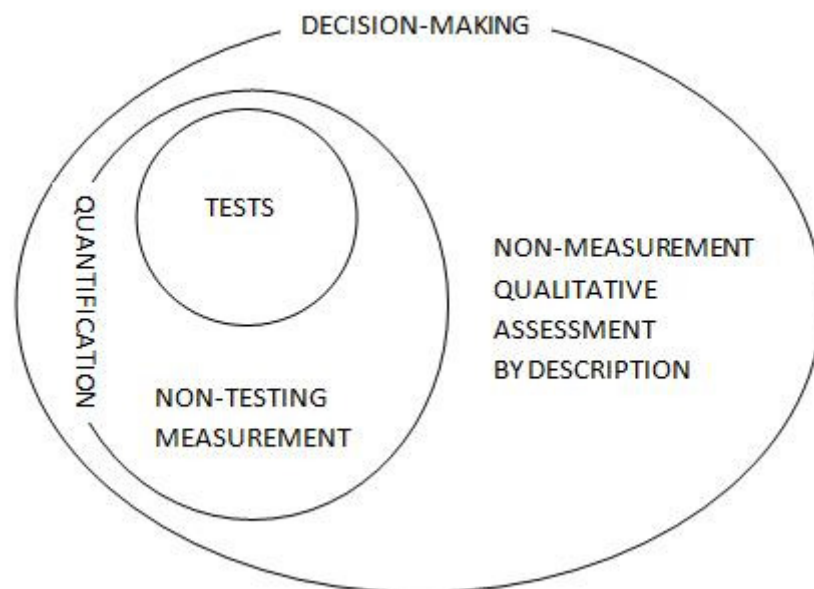

Figure 3- Alternatives in Assessment; decision-making in educational settings

As Figure 3 shows, tests constitute only a small set of options, among a wide range of other options, for a language teacher to make decisions about students. The judgment emanating from a test is not necessarily more valid or reliable from the one deriving from qualitative procedures since both should meet reliability or validity criteria to be considered as informed decisions. The area circumscribed within quantitative decision-making is relatively small and represents a specific choice made by the teacher at a particular time in the course while the vast area outside which covers all non-measurement qualitative assessment procedures represents the wider range of procedures and their general nature. This means that the qualitative approaches which result in descriptions of individuals, as contrasted to quantitative approaches which result in numbers, can go hand in hand with the teaching and learning experiences in the class and they can reveal more subtle shades of students' proficiency. This in turn can lead to more illuminating insight about future progress and attainment of goals. However, the options discussed above are not a matter of *either-or* (traditional vs. alternative assessment) rather the language teacher is free to choose *the one alternative* (among alternatives in assessment) which best suits the particular moment in his particular class for particular students.

References:

Ary, D., Jacobs, L. C. and Razavieh, A. (1996). *Introduction to Research in Education*. New York: Harcourt Brace College Publishers.

Aschbacher, P. A. (1991). Performance assessment: State activity, interest, and concerns. *Applied Measurement in Education, 4*(4), 275-288.

Bachman, L.F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bachman, L.F. and **Palmer, A.S. (**1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.

Brown, J.D. and **Hudson, T. (**1998). The alternatives in language assessment. *TESOL Quarterly* 32, 653–75.

Carroll, J. B. (1968). 'The psychology of language testing' in Davies, A. (ed.) *Language Testing Symposium. A Psycholinguistic Perspective*. London: Oxford University Press. 46-69.

Herman, J.L., Aschbacher, P.R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Huerta-Macias, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal* 5, 8–11.

Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing 18 (4)* 351–372.

McNamara, T. (2000). *Language Testing*. Oxford: Oxford University Press.

McNamara, T. (2001). Rethinking alternative assessment. *Language Testing 18 (4)* 329–332.

Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing 18 (4)* 373–391.

Weiss, C. H. (1972). *Evaluation Research: Methods for Assessing Program Effectiveness.* Englewood Cliffs, NJ: Prentice-Hall.

Biodata:

Ali Bastanfar is a lecturer in TEFL at Islamic Azad University-Khoy Branch, Khoy, Iran. He is also doing his Ph.D. in TEFL. He has more than ten years' experience of teaching English to foreign learners at various levels. His teaching experience with TEFL students at University includes teaching a vast array of courses for pre-service teachers. He has made presentations at several conferences including MICELT conference in Malaysia 2006, 4th Asia TEFL conference in Japan 2006, 5th and 6th Asia TEFL conferences in Malaysia 2007 and Indonesia 2008. He has also published articles in various journals. His major research interests are materials development, reading, testing, and learning strategies. alibastanfar@yahoo.com

---

Matching Goals to CATs

To find appropriate CAT(s) use the Student Learning Outcomes table below.

1. Reflect on your own course goals.
2. Identify the goals within the list that most closely approximate your own (5 or 6 goals is adequate).
3. Click the check boxes next to those goals.
4. Click the "Submit" button.

A chart of the goals and corresponding CAT(s) will be made which can be printed out.

Student Learning Outcomes			Classroom Assessment Techniques								
Knowledge	✔ ___	AS	CT	CM	CD	IN	PA	PO	SR	WR
Demonstrates basic knowledge of facts and terms	☐ ___	___	✘	✘	___	✘	___	✘	___	✘
Demonstrates basic knowledge of concepts and theories	☐		✘	✘	✘	✘	✘			✘
Demonstrates synthesis and integration of information and ideas	☐		✘	✘	✘	✘		✘		✘
Develops skill in using materials, tools and technology central to subject	☐					✘	✘	✘		
Learns techniques and methods used to gain new knowledge in subject	☐						✘			
Learns to evaluate methods and materials of this subject	☐				✘		✘			
Learns modeling methods appropriate for subject	☐		✘				✘			
Learns to appreciate important contributions of this subject	☐									
Develops an informed understanding of the role of science and technology	☐									

Skills

Analytical Skills	✔	AS	CT	CM	CD	IN	PA	PO	SR	WR
Analyzes problems from different points of view	☐		✘		✘	✘		✘		✘
Recognizes interrelationships among problems and issues	☐		✘	✘	✘	✘		✘		
Analyzes and interprets experimental data effectively	☐					✘	✘			✘
Applies principles and generalizations to new problems and situations	☐		✘		✘	✘	✘			
Organizes information into meaningful categories	☐			✘	✘			✘		
Uses order of magnitude estimation effectively	☐				✘					

Communication Skills	✔	AS	CT	CM	CD	IN	PA	PO	SR	WR
Communicates in writing effectively	☐		✘		✘	✘		✘		✘
Communicates in speaking effectively	☐		✘			✘				
Uses facts to get points across to others	☐		✘			✘				
Uses graphs effectively to support points being made	☐		✘	✘		✘	✘	✘		

Research Skills	✔	AS	CT	CM	CD	IN	PA	PO	SR	WR
Designs an appropriate experiment to answer a question	☐						✘			
Carries out a desgined experiment	☐						✘			
Brings in information from outside sources	☐					✘				
Uses computer-based and other resources effectively	☐					✘	✘	✘		
Seeks information on problems from multiple sources	☐					✘		✘		
Understands importance of what has already been done to solve problems	☐									
Uses appropriate synthetic/analytic methods to solve problems	☐		✘		✘	✘	✘			
Uses instrumentation appropriately and effectively	☐						✘			
Demonstrates ability to formulate effective questions	☐					✘				✘
Challenges the way things are done	☐									

		✔	AS	CT	CM	CD	IN	PA	PO	SR	WR
	Improves on what has been done before	☐		✘							
	Generates many potential solutions to a given problem	☐					✘				
Teamwork Skills		✔	AS	CT	CM	CD	IN	PA	PO	SR	WR
	Helps reconcile differences of opinion among team members	☐									
	Shares credit for success with others	☐									
	Cooperates with others	☐		✘							
	Encourages participation among all team members	☐									✘
	Shares information with others	☐		✘							
	Contributes his/her share of project workload	☐									
	Demonstrates ability to work on multidisciplinary team	☐									
	Demonstrates ability to take leadership role in support of team goals	☐									
Attitudes		✔	AS	CT	CM	CD	IN	PA	PO	SR	WR
	Identifies desirable course components	☐	✘								
	Identifies desirable course pedagogies	☐	✘								
	Identifies perceived lab/lecture match	☐	✘								
	Identifies beliefs about the nature of a field	☐	✘								
	Indicates perceptions about interdisciplinary connections	☐	✘								
	Indicates student's perceived level of understanding	☐	✘								
	Indicates student's level of confidence	☐	✘								
Instructor Goals		✔	AS	CT	CM	CD	IN	PA	PO	SR	WR
	Answers student questions on a regular basis	☐									✘
	Use assessment regularly to assure learning is occurring	☐		✘							✘
	Communicates desire for student success	☐		✘							✘
	Develops and refines instruction based on student feedback	☐		✘							✘
	Receives regular feedback from students	☐		✘							✘

Tracks typical questions	☐									✗
Retention of material	☐									
Subsequent performance in next course	☐									
Improvement in attendance	☐		✗							
Equitable performance for all students	☐									

EXPLORING RELIABILITY IN ACADEMIC ASSESSMENT

Written by Colin Phelan and Julie Wren, Graduate Assistants, UNI Office of Academic Assessment (2005-06)

Reliability is the degree to which an assessment tool produces stable and consistent results.

Types of Reliability

1. **Test-retest reliability** is a measure of reliability obtained by administering the same test twice over a period of time to a group of individuals. The scores from Time 1 and Time 2 can then be correlated in order to evaluate the test for stability over time.

   Example: A test designed to assess student learning in psychology could be given to a group of students twice, with the second administration perhaps coming a week after the first. The obtained correlation coefficient would indicate the stability of the scores.

2. **Parallel forms reliability** is a measure of reliability obtained by administering different versions of an assessment tool (both versions must contain items that probe the same construct, skill, knowledge base, etc.) to the same group of individuals. The scores from the two versions can then be correlated in order to evaluate the consistency of results across alternate versions.

   Example: If you wanted to evaluate the reliability of a critical thinking assessment, you might create a large set of items that all pertain to critical thinking and then randomly split the questions up into two sets, which would represent the parallel forms.

3. **Inter-rater reliability** is a measure of reliability used to assess the degree to which different judges or raters agree in their assessment decisions. Inter-rater reliability is useful because human observers will not necessarily interpret answers the same way; raters may disagree as to how well certain responses or material demonstrate knowledge of the construct or skill being assessed.

Example: Inter-rater reliability might be employed when different judges are evaluating the degree to which art portfolios meet certain standards. Inter-rater reliability is especially useful when judgments can be considered relatively subjective. Thus, the use of this type of reliability would probably be more likely when evaluating artwork as opposed to math problems.

4. **Internal consistency reliability** is a measure of reliability used to evaluate the degree to which different test items that probe the same construct produce similar results.

   A. **Average inter-item correlation** is a subtype of internal consistency reliability. It is obtained by taking all of the items on a test that probe the same construct (e.g., reading comprehension), determining the correlation coefficient for each *pair* of items, and finally taking the average of all of these correlation coefficients. This final step yields the average inter-item correlation.

   B. **Split-half reliability** is another subtype of internal consistency reliability. The process of obtaining split-half reliability is begun by "splitting in half" all items of a test that are intended to probe the same area of knowledge (e.g., World War II) in order to form two "sets" of items. The *entire* test is administered to a group of individuals, the total score for each "set" is computed, and finally the split-half reliability is obtained by determining the correlation between the two total "set" scores.

Validity refers to how well a test measures what it is purported to measure.

Why is it necessary?

While reliability is necessary, it alone is not sufficient. For a test to be reliable, it also needs to be valid. For example, if your scale is off by 5 lbs, it reads your weight every day with an excess of 5lbs. The scale is reliable because it consistently reports the same weight every day, but it is not valid because it adds 5lbs to your true weight. It is not a valid measure of your weight.

Types of Validity

1. Face Validity ascertains that the measure appears to be assessing the intended construct under study. The stakeholders can easily assess face validity. Although this is not a very "scientific" type of validity, it may be an essential component in enlisting motivation of stakeholders. If the stakeholders do not believe the measure is an accurate assessment of the ability, they may become disengaged with the task.

Example: If a measure of art appreciation is created all of the items should be related to the different components and types of art. If the questions are regarding historical time periods, with no reference to any artistic movement, stakeholders may not be motivated to give their best effort or invest in this measure because they do not believe it is a true assessment of art appreciation.

2. Construct Validity is used to ensure that the measure is actually measure what it is intended to measure (i.e. the construct), and not other variables. Using a panel of "experts" familiar with the construct is a way in which this type of validity can be assessed. The experts can examine the items and decide what that specific item is intended to measure. Students can be involved in this process to obtain their feedback.

Example: A women's studies program may design a cumulative assessment of learning throughout the major. The questions are written with complicated wording and phrasing. This can cause the test inadvertently becoming a test of reading comprehension, rather than a test of women's studies. It is important that the measure is actually assessing the intended construct, rather than an extraneous factor.

3. Criterion-Related Validity is used to predict future or current performance - it correlates test results with another criterion of interest.

Example: If a physics program designed a measure to assess cumulative student learning throughout the major. The new measure could be correlated with a standardized measure of ability in this discipline, such as an ETS field test or the GRE subject test. The higher the correlation between the established measure and new measure, the more faith stakeholders can have in the new assessment tool.

4. Formative Validity when applied to outcomes assessment it is used to assess how well a measure is able to provide information to help improve the program under study.

Example: When designing a rubric for history one could assess student's knowledge across the discipline. If the measure can provide information that students are lacking knowledge in a certain area, for instance the Civil Rights Movement, then that assessment tool is providing meaningful information that can be used to improve the course or program requirements.

5. Sampling Validity (similar to content validity) ensures that the measure covers the broad range of areas within the concept under study. Not everything can be covered, so items need to be sampled from all of the domains. This may need to be completed using a panel of "experts" to ensure that the content area is adequately sampled. Additionally, a panel can help limit "expert" bias (i.e. a test reflecting what an individual personally feels are the most important or relevant areas).

Example: When designing an assessment of learning in the theatre department, it would not be sufficient to only cover issues related to acting. Other areas of theatre such as lighting, sound, functions of stage managers should all be included. The assessment should reflect the content area in its entirety.

What are some ways to improve validity?
1. Make sure your goals and objectives are clearly defined and operationalized. Expectations of students should be written down.
2. Match your assessment measure to your goals and objectives. Additionally, have the test reviewed by faculty at other schools to obtain feedback from an outside party who is less invested in the instrument.
3. Get students involved; have the students look over the assessment for troublesome wording, or other difficulties.
4. If possible, compare your measure with other measures, or data that may be available.

References

American Educational Research Association, American Psychological Association, &

> National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: Authors.

Cozby, P.C. (2001). Measurement Concepts. *Methods in Behavioral Research* (7th ed.).

> California: Mayfield Publishing Company.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.). *Educational*

> *Measurement* (2nd ed.). Washington, D. C.: American Council on Education.

Moskal, B.M., & Leydens, J.A. (2000). Scoring rubric development: Validity and

> reliability. *Practical Assessment, Research & Evaluation, 7*(10). [Available online: http://pareonline.net/getvn.asp?v=7&n=10].

The Center for the Enhancement of Teaching. How to improve test reliability and

> validity: Implications for grading. [Available online: http://oct.sfsu.edu/assessment/evaluating/htmls/improve_rel_val.html].

Outcomes Assessment
Office of the Provost

Contents on This Page

I. INTRODUCTION

Academic outcomes assessment is based on a process in which faculty and staff have identified the most appropriate objectives for specific programs, e.g., general education, undergraduate and graduate majors. It employs a wide variety of measurements to discover as accurately as possible whether the department and the institution are achieving the announced objectives in these areas.

The purpose of assessment is to produce feedback to the department, school/college, or administrative unit on the performance of its curriculum, learning process, and/or services, thereby allowing each unit to improve its programs. It is not an evaluation of individual students or of individual faculty or staff.

The goal of this document is to assist chairs and other interested faculty in developing assessment plans at the departmental level. Assessment methods and instrumentation being used by academic units at UW–Madison and other comparable institutions are described here, with the intention that departments will select and/or adapt the methods best suited to their educational goals and programs. An outline of useful steps for developing a departmental assessment plan that can be used by those involved in the assessment process is also provided in this document.

II. BACKGROUND

During the last decade, colleges and universities have been called upon by a strong and influential externally driven movement to publicly demonstrate how academic programs continuously improve. National organizations and agencies, and some state legislatures, have been among those demanding more visible accountability and concrete verification that fiscal and human resources invested in educational institutions are being used in ways that result in high quality education. As one means to require accountability, many of these organizations and agencies are requesting that institutions of higher education use assessment of student learning outcomes as a means of demonstrating valuable and/or improving academic programs.

The UW System has required some form of student outcomes assessment from all UW institutions since 1900 in order to demonstrate to the Board of Regents and the legislature

that the university is being responsive to the public demand for greater accountability. In addition, the North Central Association of Colleges and Schools (NCA), UW–Madison's institutional accrediting agency, adopted student outcomes assessment as one of its requirements in 1993. It mandated that all its affiliated institutions develop a comprehensive institutional assessment plan by June, 1995 and UW–Madison has received NCA approval of its plan. NCA requires that all member institutions employ student outcomes assessment measures in general education and in all undergraduate and graduate majors.

The NCA Commission on Institutions of Higher Education identified ten characteristics of an effective program to assess student academic achievement:

1. Successful assessment flows from the institution's mission and educational purposes.

2. Successful assessment emerges from a conceptual framework.

3. Successful assessment is marked by faculty ownership and responsibility.

4. Successful assessment has institution-wide support.

5. Successful assessment relies on multiple measures.

6. Successful assessment provides feedback to students and the institution.

7. Successful assessment is cost-effective.

8. Successful assessment does not restrict or inhibit goals of access, equity, and diversity established by the institution.

9. Successful assessment leads to improvement.

10. Successful assessment includes a process for evaluating the assessment program.

NCA recognizes that faculty determination of the crucial issues of each academic program is essential and that student outcomes assessment planning is most effectively devised by faculty and staff at the departmental level. A successful program will also address the need for students to understand the purpose of assessment.

The growth of the assessment movement during the last decade has demonstrated that assessment is becoming an important tool for better understanding and responding to the needs of an increasingly diverse student population. Colleges and universities are increasingly turning to both nationally developed and locally designed assessment methods and instruments as a means of improving teaching and learning practices. The rationale for the increased focus on the development of assessment programs in academic majors is grounded in the belief that collecting systematic data improves awareness of how well students can integrate content, skills, and attitudes. Assessment research has provided useful information and insight on how students learn and what students learn, going beyond traditional measures that provide useful but limited student and programmatic data.

In responding to NCA's mandated outcomes assessment requirement, many research institutions developed assessment plans that described their institutional strategies for incorporating outcomes assessment into academic units throughout campus. These institutions vary greatly in the progress they have made in developing and implementing their respective institutional and department assessment plans. For example, because of

legislative or governing board mandates, The University of Washington, the University of Colorado, and the University of Iowa have already developed and implemented institutional assessment plans in all undergraduate majors. However, these universities remain in the early phases of incorporating outcomes assessment in graduate education. Other large research universities such as Pennsylvania State, Ohio State, Indiana, and Minnesota are in the developmental stages of assessment planning focusing significant attention on undergraduate assessment and strategies for assessing general education.

III. UW–MADISON'S ASSESSMENT APPROACH
In keeping with this institution's history of giving schools and colleges the maximum possible authority for their respective academic plans, the UW–Madison plan is constructed primarily on the basis of a plan for each college along with an institutional overview. The schools/colleges in turn have required each of their departments to develop assessment plans in order to maximize faculty involvement. Only through departmental involvement and commitment will assessment practices reflect the uniqueness of disciplines and the diversity of educational goals and missions determined by the academic unit.

IV. UW–MADISON'S ASSESSMENT PLANNING CHART
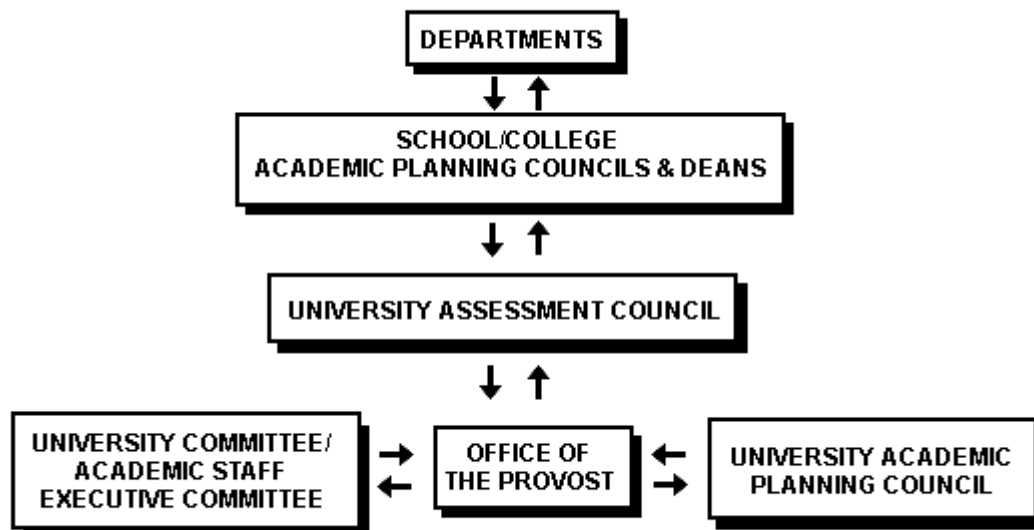
UW–Madison Assessment Plan

V. DEVELOPING A DEPARTMENTAL ASSESSMENT PLAN
When developing and implementing outcomes assessment strategies, academic units should have at least one of three purposes in mind: to improve, to inform, and/or to prove. The results from an assessment process should provide information which can be used to determine whether or not intended outcomes are being achieved and how the programs can be improved. An assessment process should also be designed to inform departmental faculty and other decision-makers about relevant issues that can impact the project and student learning.

When developing assessment programs that measure student learning to determine programmatic strengths and weaknesses, faculty often ask, "Aren't course grades a satisfactory measure of student performance?" Course grades are one source of information

about student achievement. But there are significant short-comings for basing assessment of student learning solely on course grades. A traditional letter grade may suggest how much, and perhaps how well, individual students have learned the prescribed information being tested on that particular exam, but the grades, either singly or in combination, do not necessarily reflect the role of that test in the context of the overall departmental objectives for the major. A different view, such as one or more of the suggested assessment methods, will help to focus on the overall objectives.

Developing a program-specific plan to meet assessment objectives is not an easy process. The following six step approach has enabled many academic units to develop effective plans for assessing student learning in the major.

STEP 1: Define educational/programmatic goals and objectives for the major or program.

A department's instructional goals and objectives serve as the foundation for assessment planning. Program assessment is intended to provide information on how well students are performing relative to the educational goals and objectives established by the department. The defined goals and objectives should be far-reaching and describe a variety of skills and knowledge-based areas. In most instances, not all of the goals and objectives can be adequately assessed for student achievement. However, assessment plans should be devised to assist faculty in determining whether students are acquiring some of the prescribed goals. Clearly, departmental goals for the major must ultimately be integrated with those of the school/college, which in turn, must be aligned with the institutional mission statement.

STEP 2: Identify and describe instruments or methods for assessing student achievement at important stages in the program.

Once educational goals and objectives have been identified, assessment methods for collecting student data can be chosen. These methods should be consistent with the programmatic objectives defined in the first step. Because departments often define a variety of educational goals and objectives, comprehensive assessment strategies frequently require the use of more than one assessment instrument to determine program effectiveness. (See section titled, **Assessment Instruments and Methods for Assessing Student Learning in the Major**).

STEP 3: Determine how the results will be disseminated and used for program improvement.

Assessment results and information should be used in a timely fashion to facilitate continuous programmatic improvements. Designing a feedback process is essential in all assessment plans because it gives faculty the opportunity to use recent findings to incorporate curricular changes necessary to prepare students with the skills and knowledge to advance in their respective majors. For example, when assessment results are used in a timely manner, faculty may determine that it is necessary to provide curricular changes to enhance programmatic weaknesses. When results indicate that students are performing consistently with established objectives, faculty may focus assessment initiatives in other areas or extend current practices to impact additional students.

STEP 4: Develop a timetable for accomplishing the previous three steps. Each academic unit will need to establish a schedule for selecting, implementing, and using the results of assessment strategies.

In order to meet external demands for assessment implementation and to incorporate assessment into ongoing curricular planning, departments should devise appropriate

timetables for development and execution of assessment programs. The timetables should indicate when departments foresee developing each of the previous three assessment planning steps. (For another procedure for accomplishing these four steps, see Appendix b.)
STEP 5: Submit assessment objectives, methods, and timetable to school/college Academic Planning Councils.

Each school/college will determine its specific procedures for approval of departmental plans and subsequent reviews of assessment activities. Some phases of the department's assessment plans should be carried out each academic year regardless of the frequency with which the school/college officially reviews departmental assessment activities. Departments should document all assessment activities and be prepared to demonstrate how information generated from assessment programming has been used for curricular changes by faculty.

STEP 6: Implement assessment plans and revise as needed.

Once approved by the School/College Academic Planning Councils, departments should implement assessment strategies. When initial program feedback from assessment practices becomes available, departments should use the results for programmatic improvement or to revise objectives or plans, if necessary.

By following this six step process, the complexities associated with developing effective and efficient assessment plans, especially for those devising assessment strategies for the first time, can be made less arduous and time consuming.
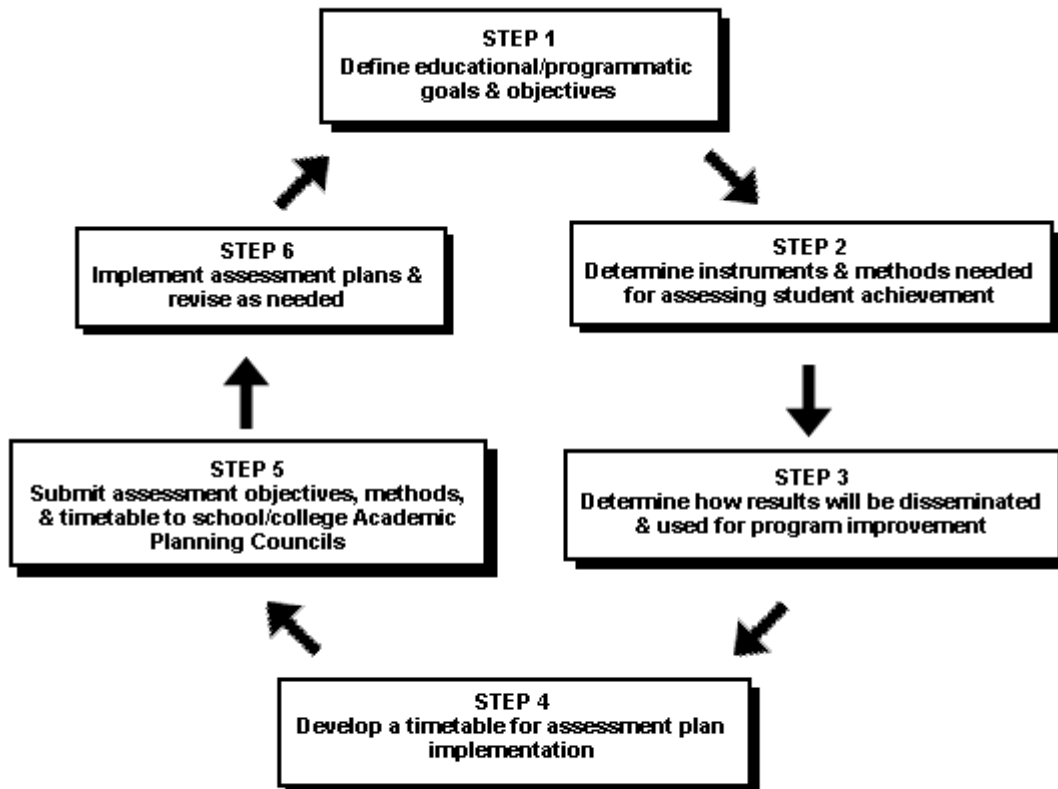
Currently departments throughout campus are in the process of reviewing and/or determining the most appropriate educational goals and objectives for each major, devising ways to measure whether students are achieving the prescribed objectives, and designing processes to employ gathered assessment data and information into curricular planning. As departments progress in the implementation of their assessment strategies, it will be important to learn from their successes and failures as they attempt to discover useful methods of measuring student achievement. The development of this manual is one effort to assist this cooperative learning effort and the Provost's Office will attempt to keep it electronically up-to-date so that it can best assist this process through shared information among colleagues at this institution and with others throughout the country.

The university has also established a University Assessment Council (UAC) to aid assessment activities. The UAC is comprised of representatives appointed by each school/college Dean, the directors of general education assessment for quantitative reasoning and communication, and representatives of each of the major assessment support services. It is chaired by a member of the Provost's staff. Its primary purpose is to share assessment ideas, particularly but not exclusively the successful efforts, in order to inform and expedite assessment efforts throughout the entire institution.

The Council believes that if properly developed and implemented, assessment of student learning in all majors can be a beneficial tool for facilitating ongoing curricular dialogue and encouraging constant programmatic improvement throughout campus. However, only through widespread faculty and departmental involvement can an institution as complex as ours devise effective and efficient program-based assessment plans that will produce results beneficial for all academic units. With assessment planning located primarily in the departments, faculty exercise their responsibility to devise appropriate methods to measure student academic achievement and program effectiveness. This process gives widespread ownership of assessment planning to faculty and enables them to determine the methods and instruments that are most applicable to their educational objectives and missions. Also,

the Council supports the idea that the academic units are best suited to determine how assessment results can be used to ascertain curricular strengths and weaknesses to improve programs.

Developing and Implementing a Departmental Assessment Plan for Programmatic Improvement

http://www.provost.wisc.edu/assessment/manual/manual2.html

University of Wisconsin-Madison

Outcomes Assessment
Office of the Provost

VI. ASSESSMENT INSTRUMENTS AND METHODS AVAILABLE TO ASSESS STUDENT LEARNING IN THE MAJOR

Assessment of student learning can be conducted using a variety of available instruments and methods. Many experts believe that a combination of assessment approaches can be the most effective way to measure student learning. Fortunately for assessment planners, many departments on campus and at other institutions have acquired some experience with many of the more commonly used instruments. Faculty in a variety of academic programs at large and small research universities have tested and used a wide range of assessment methods to determine whether students were attaining prescribed educational goals. In this section, many of these assessment approaches will be presented providing handbook users with information that can simplify the development of assessment strategies.

    A.   Direct Indicators of Learning
        1.   Capstone Course Evaluation
        2.   Course-Embedded Assessment
        3.   Tests and Examinations (Locally/Faculty Designed & Commercially Produced Standardized Tests)
        4.   Portfolio Evaluation
        5.   Pre-test/Post-test Evaluation
        6.   Thesis Evaluation
        7.   Videotape and Audiotape Evaluation of Performance
    B.   Indirect Indicators of Learning
        1.   External Reviewers
        2.   Student Surveying and Exit Interviewing
        3.   Alumni Surveying
        4.   Employer Surveying
        5.   Curriculum and Syllabus Analysis

A. Direct Indicators of Learning

1. Capstone Course Evaluation

Capstone courses integrate knowledge, concepts, and skills associated with an entire sequence of study in a program. This method of assessment is unique because the courses themselves become the instruments for assessing student teaching and learning. Evaluation of students' work in these courses is used as a means of assessing student outcomes. For academic units where a single capstone course is not feasible or desirable, a department may designate a small group of courses where competencies of completing majors will be measured.

Capstone courses provide students with a forum to combine various aspects of their programmatic experiences. For departments and faculty, the courses provide a forum to assess student achievement in a variety of knowledge and skills-based areas by integrating their educational experiences. Also, these courses can provide a final common experience for student in the discipline.

Many research universities are currently using capstone courses in a variety of academic disciplines including general education programs and other academic units in the Arts and Sciences. Departments at other research institutions using this instrument to gather information about student learning in the major include many general education programs, chemistry, political science, physics, music, religious studies, theatre, history, and foreign languages.

Relevant Publications

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

2. Julian, Faye D. "The Capstone Course as an Outcomes Tests for Majors."*Assessment in Practice*. Banta, Trudy W., Lund, Jon P., Black, Karen E., & Oblander, Frances W., (Eds). San Francisco: Jossey-Bass Publishers, 1996. pp. 79-81.

2. Course-Embedded Assessment

Assessment practices embedded in academic courses generate information about what and how students are learning within the program and classroom environment. Course-embedded assessment takes advantage of already existing curricular offerings by using standardized data instructors already collect or by introducing new assessment measures into courses. The embedded methods most commonly used involve the development and gathering of student data based on questions placed in course assignments. These questions, intended to assess student outcomes, are incorporated or embedded into final exams, research reports, and term papers in senior-level courses. The student responses are then evaluated by two or more faculty to determine whether or not the students are achieving the prescribed educational goals and objectives of the department. This assessment is a separate process from that used by the course instructor to grade the exam, report, or term paper.

There are a number of advantages to using course-embedded assessment. First, student information gathered from embedded assessment draw on accumulated educational experiences and familiarity with specific areas or disciplines. Second, embedded assessment often does not require additional time for data collection, since instruments used to produce student learning information can be derived from course assignments already planned as part of the requirements. Third, the presentation of feedback to faculty and students can occur very quickly creating a conducive environment for ongoing programmatic improvement.

Finally, course-embedded assessment is part of the curricular structure and students have a tendency to respond seriously to this method. Departments at other research institutions using embedded assessment include general education programs, classics, economics, English, film studies, geography, fine arts, history, kinesiology, philosophy, political science, physics, and religious studies.

3. Tests and Examinations

In most cases, a test will be one part of a fully developed assessment plan. Tests are commonly used in association with cognitive goals in order to review student achievement with respect to a common body of knowledge associated with a discipline or program. Departments have traditionally used tests in assessment programming to measure whether students have acquired a certain process- and content-related knowledge.

Using this approach, there are two primary testing alternatives; first, locally developed/ faculty generated tests and examinations, and (2) commercially produced standardized tests and examinations. Locally developed testing and examinations are probably the most widely used method for evaluating student progress. For assessing the validity of an academic program, examinations designed by the instructors who set the educational goals and teach the courses is often the best approach. Cost benefits, interpretation advantages, and quick turnaround time all make using locally designed tests an attractive method for assessing student learning.

Tests designed for a specific curriculum can often prove more valuable when assessing student achievement than commercial instruments. These tests focus on the missions, goals, and objectives of the departments and permit useful projections of student behavior and learning. A well-constructed and carefully administered test that is graded by two or more judges for the specific purpose of determining program strengths and weaknesses remains one of the most popular instruments for assessing most majors. Departments at other research institutions using locally designed tests and examinations include mathematics, physical education, psychology, and English.

Commercially generated tests and examinations are used to measure student competencies under controlled conditions. Tests are developed and measured nationally to determine the level of learning that students have acquired in specific fields of study. For example, nationally standardized multiple-choice tests are widely used and assist departments in determining programmatic strengths and weaknesses when compared to other programs and national data. Compilations of data on the performance of students who voluntarily take national examinations such as GRE and MCAT enable faculty to discover useful data that often leads to programmatic improvements.

When using commercially generated tests, national standards are used as comparative tools in areas such as rates of acceptance into graduate or professional school, rates of job placement, and overall achievement of students when compared to other institutions. In most cases, standardized testing is useful in demonstrating external validity.

There are a number of advantages for using commercial/standardized tests and examinations to measure student achievement; first, institutional comparisons of student learning are possible. Second, very little professional time is needed beyond faculty efforts to analyze examinations results and develop appropriate curricular changes that address the findings. Third, in most cases, nationally developed tests are devised by experts in the discipline. Fourth, tests are traditionally given to students in large numbers and do not require faculty involvement when exams are taken by students.

As part of their assessment efforts, many institutions and programs already use a multitude of commercially generated examination and tests. Some of the more commonly used national tests include:

- *ACT - COMP (College Outcome Measures Program)*: This is an assessment instrument that measures knowledge and skills acquired by students in general education courses. Administered by ACT, Iowa City, IA.
- *GRE (Graduate Record Examinations)*: The GRE is widely used by colleges, universities, departments, and graduate schools to assess verbal and quantitative student achievement. Also, many discipline-specific examinations are offered to undergraduate students in areas such as Biology, Chemistry, Education, Geology, History, Literature, Political Science, Psychology, and Sociology. The GRE is published and administered by Educational Testing Services, Princeton, New Jersey.
- *Major Field Achievements Tests*: Major field examinations are administered in a variety of disciplines. They often are given to student upon or near completion of their major field of study. These tests assess the ability of students to analyze and solve problems, understand relationships, and interpret material. Major field exams are published by Educational Testing Services, Princeton, New Jersey.

Departments with a successful history in using commercial tests and examinations include many general education programs, mathematics, chemistry, biology, computer science, geology, physics, psychology, sociology, education, engineering, foreign languages, music, exercise science, and literature.

Relevant Publications

1. Anthony, Booker T. "Assessing Writing through Common Examinations and Student Portfolios." *Assessment in Practice*. In Banta, Trudy W., Lund, Jon P., Black, Karen E., & Oblander, Frances W. (Eds.) San Francisco: Jossey-Bass Publishers, 1996. pp. 213-215.
2. Kubiszyn, Tom and Borich, G. *Educational Testing and Measurement: A Guide for Writing and Evaluating Test Items*. Minneapolis, MN. Burgess Publishing Co., 1984.
3. Popham, W. J. "Selecting Objectives and Generating Test Items for Objectives-based Tests." In Harris, C., Alkins, M., & Popham, W. J. (Eds.) *Problems in Criterion-Referenced Measurement*. University of California, Los Angeles: Center for the Study of Evaluation, 1974.
4. Priestley, Michael. *Performance Assessment in Education and Training: Alternative Techniques*. Englewood Cliffs, NJ: Educational Technology Publishers, 1992.
5. Osterlind, Steven. *Constructing Test Items*. Boston: Kluwer Academic Press, 1989.

4. Portfolio Evaluation

Portfolios used for assessment purposes are most commonly characterized by collections of student work that exhibit to the faculty and the student the student's progress and achievement in given areas. Included in the portfolio may be research papers and other process reports, multiple choice or essay examinations, self-evaluations, personal essays, journals, computational exercises and problems, case studies, audiotapes, videotapes, and short-answer quizzes. This information may be gathered from in-class or as out-of-class assignments.

Information about the students' skills, knowledge, development, quality of writing, and critical thinking can be acquired through a comprehensive collection of work samples. A student portfolio can be assembled within a course or in a sequence of courses in the major. The faculty determine what information or students' products should be collected and how these products will be used to evaluate or assess student learning. These decisions are based on the academic unit's educational goals and objectives.

Portfolio evaluation is a useful assessment tool because it allows faculty to analyze an entire scope of student work in a timely fashion. Collecting student work over time gives departments a unique opportunity to assess a students' progression in acquiring a variety of learning objectives. Using student portfolios also gives faculty the ability to determine the content and control the quality of the assessed materials.

Portfolios at other research institutions are widely used and have been a part of student outcomes assessment for a long time. Departments using portfolio evaluations include English, history, foreign languages, fine arts, theatre, dance, chemistry, communications, music, and general education programs.

Relevant Publications

1. Aubrey Forrest. *Time Will Tell: Portfolio-Assisted Assessment of General Education*. Washington, DC: AAHE Assessment Forum, 1990.
2. Belanoff, Pat & Dickson, Marcia. *Portfolios: Process and Product*. Portsmouth, NH: Boynton/Cook Publishers, 1991.
3. Black, Lendley C. "Portfolio Assessment." In Banta, Trudy & Associates (Eds.)*Making a Difference: Outcomes of a Decade of Assessment in Higher Education*. San Francisco: Jossey-Bass Publishers, 1993. pp. 139-150.
4. Jones, Carolee G. "The Portfolio as a Course Assessment Tool." *Assessment in Practice*. Banta, Trudy W., Lund, Jon P., Black, Karen E., & Oblander, Frances W. San Francisco: Jossey-Bass Publishers, 1996. pp. 285-287.
5. *Portfolio News*. Portfolio Assessment Clearing House, Encinitas, CA.

5. Pre-test/Post-test Evaluation

Pre-test/post test assessment is a method used by academic units where locally developed tests and examinations are administered at the beginning and at the end of courses or academic programs. These test results enable faculty to monitor student progression and learning throughout prescribed periods of time. The results are often useful for determining where skills and knowledge deficiencies exist and most frequently develop. Academic departments at other research institutions currently using this form of assessment to measure student learning include communications, economics, geography, linguistics, theatre, and dance.

6. Thesis Evaluation

A senior or graduate student thesis, research project, or performance paper that is structured by the department to give students an opportunity to demonstrate a mastery of an array of skills and knowledge appropriate to the major can be a useful assessment instrument. Thesis evaluation has been used effectively for program improvement in such disciplines as foreign languages, literature, and the sciences.

7. Videotape and Audiotape Evaluation

Videotapes and audiotapes have been used by faculty as a kind of pre-test/post-test assessment of student skills and knowledge. Disciplines, such as theatre, music, art, communication, and student teaching, that have experienced difficulty in using some of the

other assessment methods have had significant success in utilizing videotapes and audiotapes as assessment tools.

B. Indirect Indicators of Learning

1. External Reviewers

Peer review of academic programs is a widely accepted method for assessing curricular sequences, course development and delivery, and the effectiveness of faculty. Using external reviewers is a useful way of analyzing whether student achievement correlates appropriately with departmental goals and objectives. In numerous instances, recommendations initiated by skilled external reviewers have been instrumental in identifying program strengths and weaknesses leading to substantial curricular and structural changes and improvements.

Relevant Publications

Fong, B. *The External Examiners Approach to Assessment*. Washington, DC: Association of American Colleges. 1987.

2. Student Surveying and Exit Interviewing

Student surveying and exit interviews have become increasingly important tools for understanding the educational needs of students. When combined with other assessment instruments, many departments have successfully used surveys to produce important curricular and co-curricular information about student learning and educational experiences.

During this process, students are asked to reflect on what they have learned as majors in order to generate information for program improvement. Through using this method, universities have reported gaining insight into how students experience courses, what they like and do not like about various instructional approaches, what is important about the classroom environment that facilitates or hinders learning, and the nature of assignments that foster student learning.

In most cases, student surveys and exit interviews are conducted in tandem with a number of other assessment tools. In many universities where surveys have been adopted as a method of program assessment, findings have results in academic and service program enhancement throughout campus. Among the departments currently using these methods are general education programs, mathematics, philosophy, social work, speech and hearing science, chemistry, biology, fine arts, geology, kinesiology, and engineering.

Relevant Publications

1. Lenning, O. Use of Cognitive Measures in Assessment. In Banta, T. W. (Ed.)*Implementing Outcomes Assessment: Promise and Perils*. New Directions for Institutional Research, no. 59. San Francisco: Jossey-Bass, p. 41-52.
2. Muffo, John A., & Bunda, Mary Anne. "Attitude and Opinion Data." In Banta, Trudy & Associates (Eds.) *Making a Difference: Outcomes of a Decade of Assessment in Higher Education*. San Francisco: Jossey-Bass Publishers, 1993. pp. 139-150.
3. Riess, R. Dean, & Muffo, John A. "Exit Interviews in Mathematics." *Assessment in Practice*. Banta, Trudy W., Lund, Jon P., Black, Karen E., & Oblander, Frances W. San Francisco: Jossey-Bass Publishers, 1996. pp. 129-131.
4. Staik, Irene M., & Rogers, Julia S. "Listening to Your Students." *Assessment in Practice*. Banta, Trudy W., Lund, Jon P., Black, Karen E., & Oblander, Frances W. San Francisco: Jossey-Bass Publishers, 1996. pp. 132-134.

3. Alumni Surveying

Surveying of alumni is a useful assessment tool for generating data about student preparation for professional work , program satisfaction, and curriculum relevancy. As an assessment supplement, alumni surveying provides departments with a variety of information that can highlight program areas that need to be expanded or enhanced. In most cases, alumni surveying is an inexpensive way to gather data and for reestablishing relationships with individuals that want to help the program continually improve.

Relevant Publications

1. Converse, Jean M. & Pressler, Stanley. Survey Questions: *Handcrafting the Standardized Questionnaire*. Newbury Park. SAGE Publications. 1986.
2. Dyke, Janice Van, & Williams, George W. "Involving Graduates and Employers in Assessment of a Technology Program." In Banta, Trudy W., Lund, Jon P., Black, Karen E., & Oblander, Frances W. (Eds.) *Assessment in Practice*. San Francisco: Jossey-Bass Publishers, 1996. pp. 99-101.
3. Ewell, Peter. *Student Outcomes Questionnaires*: An Implementation Handbook. New York, NY: National Center for Higher Education Management Systems and the College Board. 1983.
4. McKenna, B. *Surveying Your Alumni: Guideline and 22 sample questionnaires*. Washington, DC: Council for Advancement and Support of Education. Contains 22 documented examples of alumni surveys successfully employed at private colleges.

4. Employer Surveying

Employer surveys can provide information about the curriculum, programs, and students that other forms of assessment cannot produce. Through surveys, departments traditionally seek employer satisfaction levels with the abilities and skills of recent graduates. Employers also assess programmatic characteristics by addressing the success of students in a continuously evolving job market. The advantages in using employer surveys include the ability to obtain external data that cannot be produced on campus, and the responses are often useful to help students discern the relevance of educational experiences and programs.

Relevant Publications

1. Converse, Jean M. & Pressler, Stanley. Survey Questions: *Handcrafting the Standardized Questionnaire*. Newbury Park. SAGE Publications. 1986.
2. Dyke, Janice Van, & Williams, George W. Involving Graduates and Employers in Assessment of a Technology Program.@ In Banta, Trudy W., Lund, Jon P., Black, Karen E., & Oblander, Frances W. (Eds.) *Assessment in Practice* San Francisco: Jossey-Bass Publishers, 1996. pp. 99-101.

5. Curriculum and Syllabus Analysis

In a perfect planning/implementation cycle, once a department has defined its objectives, all phases of the curriculum and each individual course would almost automatically cover all the bases needed to provide each student the opportunity to learn the essential components of those objectives. It doesn't happen that way, however, because departmental personnel change over the years and the higher education tradition of freedom within the classroom often leaves course content almost totally to individual instructors.

In any case, not every course needs to attempt to cover all the objectives for the major. As one technique to keep a focus on the agreed-upon objectives, curriculum analysis provides a means to chart just which courses will cover which objectives. The chart then provides assurance to the department that, assuming certain sequences are taken by the student candidates for that major, they will in fact have the opportunity to learn those objectives.

Syllabus analysis is an especially useful technique when multiple sections of a department course are offered by a variety of instructors. It provides assurance that each section will cover essential points without prescribing the specific teaching methods to be used in helping the students learn those objectives.

http://www.provost.wisc.edu/assessment/manual/manual4.html#appen_b

VII. UW–MADISON CAMPUS-BASED ASSESSMENT RESOURCES

A number of campus-based organizations and offices have been instrumental in assisting departments in developing appropriate assessment programs and methodologies for assessing student learning in the major. They include:

➢ Testing and Evaluation (T&E)

➢ University of Wisconsin Survey Center

➢ Office of Quality Improvement (OQI)

➢ Office of the Provost

Testing and Evaluation (T&E) has worked with numerous departments including Educational Psychology, French and Italian, German, Spanish and Portuguese, Mathematics, and Electrical Engineering to devise tests and methods for collecting essential data used for analyzing the effectiveness of courses and student progression. Contact Allan Cohen for more information at (608) 262-5863.

The **University of Wisconsin Survey Center (UWSC)** has been very active in assisting departments in developing student surveys to measure student achievement and satisfaction. Since 1993 UWSC has worked with faculty and administrators to conduct a campus-wide survey of undergraduate student satisfaction. During the forthcoming year, UWSC in planning on conducting a pilot alumni survey that will provide valuable information for preliminary discussions about developing a larger, more comprehensive alumni survey. Contact James Sweet, Director, for more information at (608) 262-2182.

The **Office of Quality Improvement (OQI)** assists departments with designing assessment processes and particularly to help faculty identify desired student learning outcomes. OBI stresses using assessment processes as a tool to link student learning to future strategic and curricular planning. Contact Kathleen Paris for more information at (608) 263-6856.

The **Office of the Provost** works collaboratively with the University Assessment Council to provide assistance to departments developing assessment strategies. In the Office of the Provost, discipline-specific assessment data are kept as a resource for faculty inquiring about assessment techniques, costs, and benefits from other comparable departments and institutions. Also, based on proposals submitted by representatives of the University

Assessment Council, the Office of the Provost has made available financial assistance to those departments in the developmental stages of assessment. To apply for these resources, departments submit assessment proposals to their School or College representative on the University Assessment Council. For questions, or additional information, contact Mo Noonan Bischof in the Provost's Office at (608) 262-5246 or atmabischof@wisc.edu.

VIII. CONCLUSIONS

Outcomes assessment is a valuable and integral part of programmatic improvement and review. It has the potential to impact students and faculty members in academic units throughout campus. As programs devise and incorporate assessment practices into on-going curricular structures, faculty will acquire useful information about student learning that may support existing educational practices or demonstrate that necessary changes need to occur.

In order for assessment plans to be effective, faculty must work collaboratively to develop strategies that fit with the educational missions, goals, and objectives of the department. There are no simple approaches to developing effective and efficient assessment plans. Reliable assessment programs often take years to perfect and to begin producing the type of results anticipated by committed faculty. It is hoped that the techniques outlined in this manual will be of assistance as staff of this institution implement a successful assessment program.

Appendix A

UNIVERSITY ASSESSMENT COUNCIL

Appendix B

Developing an Assessment Plan in the Major*

1. Agree on your mission

2. Create goals for student outcomes and processes

3. Identify related activities for each goal

4. Brainstorm appropriate measures

5. Evaluate and select measures

6. Identify appropriate assessment methods

7. Develop a plan for collecting data

8. Prioritize goals

9. Set timeline, milestones

10. Implement assessment plan

11. Use data to improve processes

12. Communicate results

* From Hatfield, Susan, "Assessment in the Major - Tools and Tips for Getting Started." Paper presented at the 1997 Assessment Conference in Indianapolis. Professor Hatfield is the Assessment Coordinator at Winona State University.