A Guide for Evaluating Theological Learning

Introduction

This guide is written for theological educators, and it follows closely the Standards of Accreditation of the Commission on Accrediting, particularly as they relate to the assessment of educational effectiveness. Assessment and evaluation comprise a fundamental theme throughout the general institutional and degree program standards. Each degree program standard requires that the school "shall be able to demonstrate the extent to which students have met the various goals of the degree program" (i.e., A.5.1), and this expectation follows the general model of evaluation prescribed in Standard 1: "(1) the identification of desired goals or outcomes for an educational program…; (2) a system of gathering quantitative or qualitative information related to the desired goals; (3) the assessment of the performance of the program…; and (4) the establishment of revised goals or activities based on the assessment" (1.2.2). Effective assessment of learning thus entails an ongoing process in the life of the school rather than episodic occurrences. While the Standards of Accreditation require evaluation across the broad range of institutional efforts, assessment of student learning is an increasingly important focus in higher education accreditation, and this section of the *Handbook of Accreditation* focuses on a range of issues and strategies for the assessment of learning for religious vocation.

Evaluating Theological Learning was commissioned as part of the ATS Character and Assessment of Learning for Religious Vocation project. It was written by John Harris of Samford University, and it reflects his counsel, his informed interest in theological educa-

tion, his teaching in a theological school, and a career of work in higher education that has focused on the assessment of learning, improvement of educational quality, and higher education accreditation. The Board of Commissioners adopted this text as its counsel regarding the Commission's standards related to the assessment of learning. The focus throughout this guide is on the Master of Divinity (MDiv) program, since that program is offered by virtually all Commission-accredited schools, but the perceptions of assessment are applicable to assessing learning in all the degree programs approved as part of accreditation by the Commission on Accrediting.

It has been written as a practical guide—but not a "cookbook" for assessing learning. It deals with theoretical and technical issues related to assessing higher learning because Commission-accredited schools cannot assess the sophisticated kind of learning that characterizes the best of theological learning with simplistic models of assessment. This guide is not a comprehensive resource. It does not focus on all the kinds of evaluation that the Standards of Accreditation require. Its focus is the assessment of the outcomes of learning. There are many comprehensive and technical treatises on assessment in higher education. The appendix, abbreviated from a much longer annotated bibliography developed for the 2004 Assessment Institute at Indiana University Purdue University Indianapolis, summarizes six of these. Other books on assessment have been summarized in other ATS publications,[1] and these are helpful guides, as well. A number of useful Web sites have been developed on assessment, including online assessment resources available on the ATS Web site.

The guide has six parts. The first and second parts deal with definitions and perspectives. Part 1, Assessing Outcomes in the Master of Divinity Program, focuses on the nature of outcomes in theological learning, particularly as they relate to the MDiv program. Outcomes in higher education are complex, and assessing them requires sophisticated intellectual effort. This section also

summarizes the fourfold approach to evaluation in the ATS standards. Part 2, Evaluating Institutional Characteristics in Support of Learning, explores the way in which the traditional focus of accreditation on resources relates to the emerging dominant focus on outcomes of learning. The final four sections explore principles and practices related to each of the four steps in the overall evaluation process as defined in Standard 1. These steps include: the identification of desired goals or outcomes for an educational program (explored in Part 3); the development or adoption of a system of gathering quantitative or qualitative information related to the desired goals (addressed in Part 4); the assessment of the results of information regarding learning outcomes (part 5 of this guide); and the establishment of revised goals or activities based on the assessment (part 6). The guide concludes with a summary reflection about assessment and accreditation.

The Commission on Accrediting's approach to assessment is as a process of institutional and educational stewardship. Assessment should add value, enhance educational practice, and contribute to the increased attainment of a theological school's educational mission. Assessment of theological learning adds work, to be sure, and it adds work that schools have not done previously, but it has the great potential of adding value. *Assessment adds value when it enhances quality, tracks educational attainment, and prompts educational improvement.* Accreditation evaluation committees, however, often sense that an institution's assessment program is a Potemkin Village[2]—quickly constructed before the visit and quickly dismantled following it. Enormous amounts of time and energy can go into administering and analyzing surveys and other assessment data without much evidence that the resulting information is understood or used to improve instructional practices. Assessment can ensure educational achievement and contribute to educational improvement in theological schools only as it is thoughtfully conducted, over time, with relevance to the unique learning and formational goals of theological education.

At the beginning of the twentieth century, E. L. Thorndike asserted that if something exists it can be measured. Psychological measurement and most current educational assessment theory rest on this positivist assumption. While quantitative measures continue to be important and valuable indicators, an emerging literature casts suspicion on the assessment community's over-confidence in and reliance upon numbers. Although numbers can tell us much, they do not express all of reality, including critical outcomes of theological learning.[3] However, theological educators' worry about the assessment community's over-confidence in objective assessment methods does not mean that assessment should be avoided in theological education as a flawed practice. There are other ways of understanding assessment. Michael Polanyi explained that the purpose of his *Personal Knowledge* was "to show that complete objectivity as is usually attributed to the exact sciences is a delusion and in fact is a false idea."[4] Michael Power warns about blindly adopting the audit model to ensure quality in education.[5] David Boyle, an economist, questions the belief that quantitative audits can tell us the whole truth.[6] In *Discerning Is More Than Counting*, John Harris and Dennis Sansom wrote that "We have to find our way between the Charybdis of knowing only through numbers and technique, and the Scylla of knowing only by sensing."[7] Theological educators need models of assessment that use numbers when most helpful and use qualitative, non-numeric approaches when they are most appropriate. This guide assumes a wide range of models of assessment that, it is hoped, will help seminary faculty and deans find useful paths between over-dependence on numbers and rejection of educational assessment on the basis of faulty philosophical assumptions.

1. Assessing Outcomes in the Master of Divinity Program

The Master of Divinity degree program standard requires that students be educated in four areas: (1) Religious Heritage, (2) Cultural Context, (3) Personal and Spiritual Formation, and (4)

Capacity for Ministerial and Public Leadership.[8] Because the purpose of the MDiv is to prepare individuals for ministry, each theological school is expected to ground its program in the theology and ministerial contexts of the faith communities that its graduates will serve. By mandating these four broad educational areas, ATS is not suggesting that seminaries develop the same outcomes, or that schools give the same weight to the four areas. A seminary identified with one denomination will have distinct emphases within some or all of these four content areas that differ from a seminary identified with another denomination, or from a seminary whose graduates serve many denominations or faith communities.

MDiv programs, unlike other post-baccalaureate, professional master's degree programs, differ in content and emphasis by the constituencies they serve. For example, MBA programs and other professional master's programs are more similar in content across schools than are MDiv programs. While these four areas of the MDiv mandate attention, some ATS schools give priority to spiritual and personal formation, some focus on biblical studies and languages, and still others emphasize theological construction. The MDiv standard requires each school to address the four areas, but it grants considerable latitude in determining the content within each area and the relative weight assigned to the different areas.

Assessing learning in the MDiv program depends on the identification of learning goals in the four educational areas and the four steps for assessing them identified in Standard 1. Within these four broad categories, the ATS MDiv standard requires student attainment in particular areas as summarized in Figure 1 in the column headed "MDiv Program Content."

The second column, "Approach and Outcomes," represents the need for each Commission-accredited institution to identify its particular way of dealing with the required content areas and to

Figure 1
CHECKLIST: Designing MDiv Program and Assessing Outcomes

MDiv PROGRAM CONTENT	APPROACH & OUTCOMES	ASSESSMENT STRATEGY
1. Religious Heritage		
1a. Scripture		
1b. Faith Community's Doctrine/Theology		
1c. Faith Community's History—Social/Institutional		
1d. General Christian History & Heritage		
2. Cultural Context		
2a. Contemporary social cultural issues		
2b. Global, Multicultural, Cross-cultural Church		
3. Personal & Spiritual Formation		
3a. Counseling & Reflection		
3b. Christian Faith Development		
4. Capacity for Min. & Public Leadership		
4a. Ministry Practice		
4b. Relation of practical ministry to other subjects		

define the learning outcomes it expects of students. The third column, "Assessment Strategy," identifies the school's need to develop and implement strategies for assessing student achievement in each content area. The MDiv standard dictates the content of the first column, but each school must identify the content for the other two columns in the context of its particular implementation of the MDiv. Once schools have identified the desired outcomes and strategy for assessing these outcomes, they can begin to think about comprehensive educational assessment, using the fourfold evaluation strategy.

Figure 2 shows the relationship between the learning goals in the four required areas of the MDiv program and the evaluation process described in Standard 1. The evaluation process for each of these content areas requires the identification of goals, collection of information on which reasonable decisions could be based regarding the attainment of the goals, the determination of the way in which the learning goals have been attained, followed by decisions about the value of the learning goals in light of students'

attaining or not attaining them (as shown in the row headings in Figure 2).

The attainment of learning goals in a professional degree program like the MDiv cannot be fully determined while students are in the degree program, or even at the time of graduation. MDiv students are being educated for performance in ministry, and some learning goals cannot be assessed until graduates are in ministry settings. So, Figure 2 adds a third dimension—time. There are some goals that can be assessed while in a degree program, some goals at the end of the program, and some goals only after the degree has been awarded and the graduate is in practice in the field.

Figure 2
Education and Formation Outcomes and Evaluation Steps

Thus far, the learning goals related to the four required areas of the MDiv program have been discussed as if they were discrete and unconnected. However, the standard indicates that achievement and formation in these four areas should be integrated: "Instruction in these areas shall be conducted so as to indicate their interdependence with each other and with other areas of the curriculum, and their significance for the exercise of pastoral leadership." (A.3.1.1.3)

The specific intent of the standard is to avoid construing religious heritage, cultural context, personal and spiritual formation, and capacity for ministerial and public leadership as separate silos of MDiv learning. This expectation requires that the focus for assessment be the total MDiv program, not individual course outcomes. Integrated outcomes result from an integrated curriculum and instructional strategies. Figure 3, Integrated Learning & Outcomes, illustrates this emphasis. Therefore, comprehensive assessment of learning in the MDiv program should indicate how well students have integrated Theological Heritage, Cultural Context, Personal and Spiritual Formation, and Capacity for Ministerial and Public Leadership to minister faithfully and effectively.

Figure 3
Integrated Learning & Outcomes
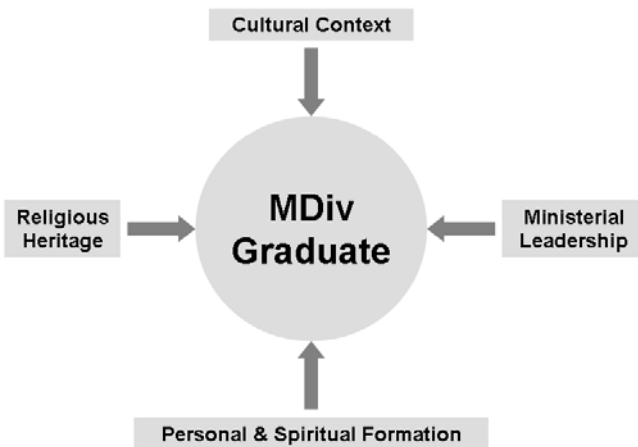
While direct evidence of students reaching stated goals is needed, this expectation does not require certification that *each* graduate reaches a predetermined proficiency in all areas. It does mean that schools should state expected student outcomes and document that graduates *generally* attain them. This key and sensitive point warrants reiteration: ATS policies and standards do *not* require that *each* graduate must reach predetermined levels of proficiency or attainment, but they do require evidence that a graduating class collectively possesses expected outcomes.

2. Evaluating Institutional Characteristics in Support of Learning

While this guide focuses on the assessment of student learning, the standards also require evaluation of institutional characteristics (such as faculty qualifications, library holdings, and financial resources). Although schools have grown accustomed to assessing these characteristics, they are at best indirect indicators of educational effectiveness. Increasingly, accreditation is requiring the use of more direct indicators of the achievement of student learning.

Because faculty qualifications, institutional facilities and finances, the library, and instructional technology s*upport* the educational and formational efforts, institutional attributes are more appropriately evaluated by their *support* of the learning experience. In this role, they may be judged by capacity, continuity, and integrity.

Capacity: Does the seminary have the institutional attributes and resources to support its educational program to reach its learning goals?
Continuity: Given its organization and resources, is the seminary likely to continue to exist and to operate effectively and faithfully?
Integrity: Is the seminary organizationally integrated and transparent?

Figure 4, Institutional Attributes & MDiv Outcomes, illustrates the role of institutional attributes in supporting the educational program rather than serving as direct assurances of learning. Their value lies in their contribution to ensuring capacity, continuity, and integrity in the educational and formational program. Increasingly, institutional characteristics will be evaluated in terms of their contribution to educational effectiveness rather than as indirect indicators of educational effectiveness.
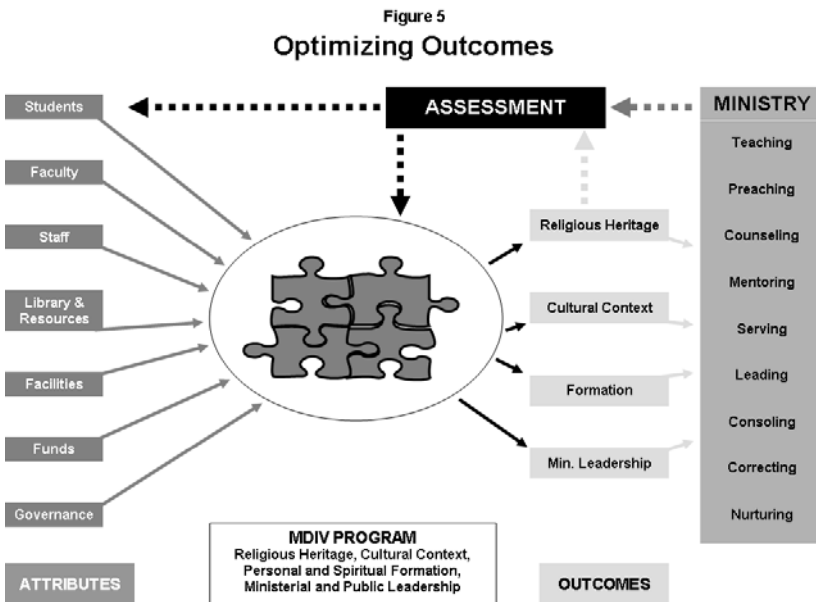
Figure 4
Institutional Attributes & MDiv Outcomes
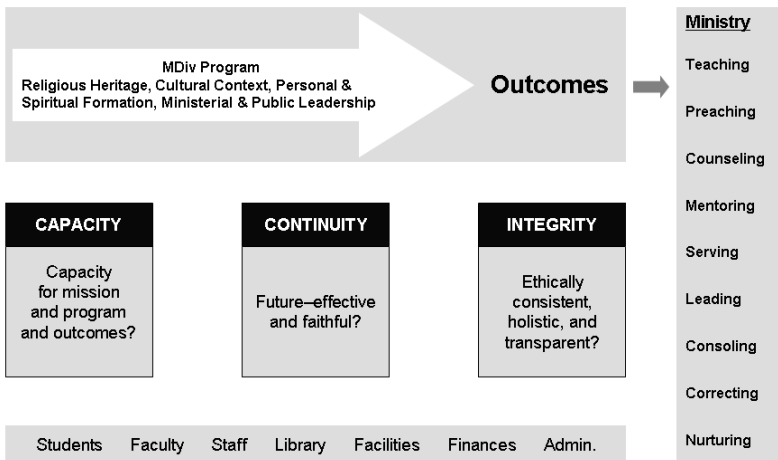
Optimizing Outcomes

Figure 5, Optimizing Outcomes, illustrates how assessment of outcomes could be used to improve MDiv education. The institutional attributes and resources (identified in Figure 4 as the foundation of an MDiv program) appear on the left in Figure 5. They comprise the resources brought together in the instructional/ formational program to yield desired outcomes in the four areas of MDiv learning. The dotted-line arrow from ministry to student outcomes suggests that assessing performance after graduation should be used in evaluating student outcomes. Similarly, assess-

ment of learning outcomes before graduation should be used to evaluate the quality of institutional characteristics that support the educational and formational program. Assessment has its greatest effect when a faculty and administration understand and visualize the MDiv program as a flow-through system. All too often, academic institutions are organized in loosely coupled, disciplinary silos that emphasize student knowledge and faculty scholarship within discipline-specific compartments. The silo organizational structure undermines the assessment of integrated student outcomes.

Figure 5
Optimizing Outcomes

In Figure 5, feedback from before-graduation assessments and after-graduation reports from alumni/ae are used to optimize institutional attributes for best results. A seminary committed to using outcomes assessment to influence its selection of faculty, curricular strategies, and learning resources will likely differ from a school focused on faculty scholarship and discipline-specific learning. An educational system that pays careful attention to these indicators of integrated learning can continuously redesign

itself to increase its educational effectiveness. When student learning consistently falls below desired goals, a search for root cause(s) begins. Questions such as these aid the search for root cause:

- Do the typical entering students have the abilities and inclinations to develop the desired characteristics of mind and heart with good instruction and formation?
- Is the curricular content optimally balanced for the desired outcomes?
- Are modes of learning and formation optimal for the desired outcomes?
- Does life in the seminary's community foster and reinforce the desired values, beliefs, commitments, and work habits?

A seminary organized as suggested in Figure 5 assumes that its effectiveness is determined by evidence of student learning and graduate's ministerial performance. Accordingly, the quality of faculty, library holdings, facilities, or financial resources is determined by their contribution to desired outcomes.

3.   Identifying Goals or Outcomes

The MDiv standard identifies areas of education, but it is the institution's responsibility to identify educational goals related to these four instructional areas.

Identifying Outcomes with a Realistic Perspective

When taken seriously, stating learning goals for a program is challenging. Faculty often have a better idea of the courses that should compose a curriculum than of the overall learning that should result from it, especially when learning goals need to be expressed in specific and observable terms. Generally, faculty need some experience with a curriculum before it can realistically identify what students should know and be able to do at the completion of the program, just as it is easier to state the goals for

a course that has been taught several times. Experience over time, with a course or an entire curriculum, generates more realistic goals.

Stating goals in terms of how current, typical graduating students differ from the typical, entering students is often better than beginning *de novo*. When a faculty is asked to articulate desired outcomes, it is easy for them to hope for more than is attainable, given the school's circumstances and resources. One way to counter this tendency toward idealization is to learn as much as possible about the knowledge, understanding, skills, values, beliefs, and habits of the current senior students. On the one hand, a reasonably accurate understanding of the actual outcomes of an MDiv education, i.e., the differences between graduating and entering students, does not define what an *ideal* MDiv graduate should know and be. On the other hand, starting with actual outcomes is often more helpful than beginning with abstract ideals. A reliable estimate of actual outcomes allows a faculty to identify hoped for outcomes.

Generally, curricula are constructed on the basis of faculty members' own educational experience and their perceptions of common practices. Rarely are they constructed from clearly identified destination points or learning goals. Because accreditation is increasingly focusing on student learning, faculties need to develop more skill in constructing curricula on the basis of desired learning goals.

Suggestion 1. In attempting to state real and useful learning goals for an existing MDiv program, consider developing a common understanding among the seminary's faculty and administration of the current, actual outcomes.

Suggestion 2. Given an understanding of current outcomes, develop desired and realistic goals.

Giving Attention to Levels of Learning

Theological education requires more than factual learning. MDiv graduates should be formed intellectually and affectively so that they can and will act on their beliefs and understanding. Such aspirations are often reflected in stated learning and formation goals. Knowing content and being able to act on it represent two different levels of learning. More valid and reliable assessments can be developed from goals that differentiate desired levels of learning. While Benjamin Bloom's taxonomies are widely known and used by instructional designers and measurement specialists, the taxonomy developed by Lee Shulman might be more useful to seminary faculty and administrators. Shulman's six levels of learning[9] do not separate cognition, affect, and performance. For Shulman, growth in knowledge, understanding, judgment, and commitment are captured in one term—learning—and for him, learning involves all of one's being. Learning, for Shulman, includes both cognitive and affective shaping of the person. This model provides a way to deal with the varying emphases in theological education in different faith communities and their seminaries.

Shulman describes each level with precision, and the following provides a working summary of each level, without the precision or elegance of Shulman's original statement.

Engagement and Motivation. Engagement can be both a proxy indicator as well as an end of learning. "Proxy indicator" suggests that the connection between engagement and learning is so strong that a measure of engagement may substitute for an actual test of the learning. But engagement can be more than an indirect assessment of learning; it can be a learning outcome. As an end, engagement signifies that the student has learned to like and pursue learning. A hoped-for outcome of theological education is that MDiv graduates found learning so meaningful that they continue to learn in and from their ministry. Documentation that graduates

become lifelong, self-starting, active learners would be an indicator of an educational program's success.

Knowledge and Understanding. Understanding is knowledge that has become owned by the learner. If knowledge is owned by learners, they can explain it in their own words. Understanding is knowledge that connects with the learner's experiences. For example, Dennis Sansom presses his students to distinguish among utilitarian, duty-based, and Christian ethics. It is one thing to know the difference, and another to understand the difference so that implications and examples of each can be explained in the student's own words.[10] Shulman observes that educators devote most of their time and effort to increasing understanding in the hope that understanding will shape behavior.

Performance and Action. Professional learning leads to performance based on understanding that is consistent with certain standards. Professional education produces skill-sets through which the needs of others are met. We assume such skills are nested in theory and that their proper use is shaped by understanding. Shulman hastens to add that, in many cases, people act in order to understand, rather than understand to act. A student's learning of the theological dimensions of leadership is not ultimately gauged by knowledge of the concept, but by the ability to exercise leadership that reflects those theological dimensions.

Reflection and Critique. Beyond the ability to act on understanding lies the ability to learn from action. Professionals learn by critiquing their work. Such reflection not only allows them to improve their work, it also allows them to question the truthfulness or validity of the understanding that shaped it. For example, a new minister in an inner city church might have an informed idea about how people from different socio-economic backgrounds and races can work together and develop a strategy consistent with that idea. If the strategy does not work, the minister should not

only question the plan but also the idea that shaped the plan. The ability to reflect on experience drives continuous learning.

Judgment and Design. Through judgment, learners adjust a general understanding to differing circumstances and realties. Judgment drives design because design takes into account certain constraints. Designing a church's educational program requires weaving together theological and educational understanding while taking account of a particular congregation's circumstances and constraints.

Commitment and Identity. At this highest level, people become what they understand, perform, critique, and evaluate. Understandings, beliefs, and inclinations consolidate into certain patterns of behavior, even moral standards. At the end of an MDiv program, it is the hope that students have learned to engage with learning, internalize and integrate understandings, act on their understandings, learn from experience, adapt to circumstances and constraints, and commit themselves to their communities of faith with integrity and courage.

Shulman's six levels of learning may be particularly useful to theological educators as they develop goals or outcome statements for MDiv programs. Figure 6 illustrates how these levels of learning can be related to the four areas of learning identified in the MDiv standard. Some theological educators may perceive these levels of learning as an implied theological anthropology inconsistent with their theological commitments. Nevertheless, these levels can serve as effective markers of growth in learning and formation. The matrix in Figure 6 should not be interpreted as requiring that any or all of the six learning levels must be addressed in the MDiv areas of learning.

Figure 6 can help in several ways. In determining goals, for example, it can remind a faculty to consider the expected levels of learning or formation. Hopefully, it will stimulate thinking about

Figure 6
MDiv Program Areas and Shulman's Levels of Learning

Shulman's Levels of Learning	MDiv Program Areas			
	Religious Heritage	Cultural Context	Spiritual Formation	Ministry and Public Leadership
Engagement and Motivation				
Knowledge and Understanding				
Performance and Action				
Reflection and Critique				
Judgment and Design				
Commitment and Identity				

the comprehensive goals of the MDiv, including both experiential and academic.

Identifying Outcomes in Discernible Terms

Evaluation is aided when goals are stated in discernible terms. The Oxford English Dictionary defines "discern" as, "to distinguish (one thing or fact) by the intellect; to recognize or perceive distinctly."[11] At its root, discernment means "to separate (things or one thing *from* another) as distinct; to distinguish and divide."[12]

Some aspects of ministerial education can be counted or measured, such as the percentage of correct answers on a Hebrew vocabulary test, the correct sequencing of biblical events on a timeline, or the identification of Roman emperors and Palestinian rulers in the first century. Many of the broader curricular outcomes of the MDiv are not amenable to quantitative measurement. Their discernment usually relies on the tacit knowledge of multiple observers rather than observer-independent, objective measurement. Nevertheless, goals can be written to facilitate consistency of discernment. Explicit and specific statements contribute to common understanding. Observers can more readily discern expected outcomes

if the outcomes refer to examples of "unacceptable," "acceptable," and "outstanding" work by former students. In addition to examples, the general statements of goals or outcomes might be supplemented by more specific operational statements. Operational statements of goals or outcomes usually contain these components.

1. Introductory phrase that describes what the student will be given to respond to.
2. Active, transitive verb that describes what the student is to do.
3. Direct object that describes the desired result.
4. Modifiers that indicate the expected level or quality of the work.

In the area of Ministerial Leadership, a seminary might expect a senior to demonstrate an accurate understanding of the polity of the student's ecclesial body. Such an understanding might be worded as follows: When presented with a case in which a lay leader is accused of unethical or immoral behavior, describe the steps that should be followed consistent with the polity of your faith community.

Understanding as a Learning Goal

A review of several seminaries' self-studies for ATS indicates that "understanding" is a frequent learning goal. One way that learning for understanding can be demonstrated is through asking students to apply or transfer a concept learned from one setting to another. For example, a student could demonstrate understanding of exegetical principles by using them with a text that has not previously been addressed in class. If understanding is a goal, the statement of the goal should indicate how that understanding can be discerned by observing students' attempts to apply it to new and different circumstances.

Elliot Eisner, with his substantive involvement in art education, distinguishes between "instructional objectives" and "expressive

objectives."[13] An instructional objective describes the specific behavior or performance that should result from an instructional program. The following statements illustrate instructional objectives:

- Given a list of Greek nouns, pronouns, and adjectives, correctly pronounce, spell, and compose sentences using all of their declensions.
- Given brief descriptions of various Greek philosophical positions, identify the person(s) considered the principal architect and proponent of each.
- Given an unorganized list of clerical positions in the Greek Orthodox Church and Roman Catholic Church, sort the positions into two lists—Greek Orthodox and Roman Catholic—and arrange each list in hierarchical order.

Although any field requires certain specific knowledge and skills, advanced professional education goes beyond this kind of knowledge. Eisner distinguishes expressive learning objectives from instructional objectives as follows:

An expressive objective does not specify the behavior the student is to acquire. . . . An expressive objective specifies an educational encounter in which [students are] to work, a problem with which they are to cope, a task in which they are to engage; but it does not specify what from that encounter, situation, problem, or task they are to learn. . . . An expressive objective is evocative rather than prescriptive.

. . . the mode of evaluation is similar to aesthetic criticism; that is, the critic appraises a product, examines its qualities and import, but does not direct the artist toward the painting of a specific type of picture. The critic's subject-matter is the work done—he does not prescribe a blueprint of its construction.

…instructional objectives emphasize the acquisition of the known; while expressive objectives its elaboration, modification, and, at times, the production of the utterly new.[14]

The expressive objective approach offers a less mechanical way to assess learning for understanding. The following examples are learning goals that involve understanding and the "expressive" ways by which that understanding can be assessed.

Understanding Goal 1. MDiv students will understand Christian moral and ethical principles so they can apply them to moral and ethical dilemmas within the social and economic context of their ministry.

Expressive Assessment for Goal 1. A young business owner in your church, whose company has just exploded from a three-person operation to a 250-employee operation, asks you to assist in developing an employee handbook and pay scale to reflect theological principles. The learner will be able to identify key principles and show how they may be translated into policies regarding the payment of the company's workforce

Understanding Goal 2. MDiv students will understand the religious heritage of their faith community so they can explain it in straightforward language.

Expressive Assessment for Goal 2. A new member of your church with only a high school education and no background in your denomination asks you to explain the distinctive doctrines and practices of your denomination. Your task is to explain some distinctive doctrines. Your response should use the least number of specialized theological terms while maintaining the integrity of your denomination's creed and polity.

Understanding Goal 3. MDiv students will grasp the distinctive demands of Christian love and how it should shape their own values, dispositions, and behaviors.

Expressive Assessment for Goal 3. Given biblical characteristics of self-giving love, construct a systematic examination of yourself in light of these characteristics, noting how your behavior patterns are consistent or inconsistent with the characteristic of self-giving love. Each student should ask two close friends to use the same technique to rate her or him.

Suggestion 3. When formulating MDiv learning and formation goals, they are more likely to be discernible if written as follows: (1) Identify to what the student will be asked to respond. (2) Use active verbs that require actions from which multiple observers can infer interior understandings and values. (3) Use direct objects that indicate what should result from the student's actions. (4) Identify conditions and criteria of performance.

4.   Collecting Quantitative or Qualitative Information Related to the Desired Goals

Using the word "collecting" instead of "testing" or "surveying" in a discussion of assessment may surprise the reader. In *Technopoly*, Neil Postman reports that William Farish, a Cambridge tutor and professor of engineering, was the first person known to assign numbers as grades on student papers.[15] Assessment typically has used tests and surveys that yield numbers. "Collect" is a much broader and less precise concept than "test" or "survey." It does not suggest special intervention as is the case with a test or survey. Rather it suggests using what is available. It includes what may be gleaned from observation of existing work and behavior. "Information" unlike "data" suggests collecting any evidence useful for estimating what students know. "Collecting information," therefore, includes gathering artifacts of student work and observations of student behaviors as well as having students take special tests or respond to surveys.

21

Using Assessment Artifacts

Before adding additional tests and questionnaires, existing student work can be used to make inferences about learning. Student papers or any student work product can serve as an artifact for assessment through which patterns of understanding and ways of thinking may be discerned. For example, final examinations and/or papers from a required first semester course might be compared to examinations and papers most students take in their last year. Thus, major pieces of student work completed for one purpose such as meeting the requirements of a course can be used as a basis for assessing overall patterns of student growth. Possible artifact examples follow:

- Tests or exams normally given in courses taken by all or most students
- Comprehensive papers or projects often done the last year
- Class or course papers or projects
- Books students select for personal reading, non-school related reading
- Performance on ordination reviews and examinations
- Observations of teamwork and mutual assistance
- How financial obligations are met
- Post-graduation career paths of graduates

While there is usually no shortage of these kinds of artifacts, faculties have not been encouraged to use them in assessment. That is, faculties have not been encouraged to use student work done for one purpose, such as to meet a course requirement, for another purpose, such as evaluating general student growth. As in excavating for archeological artifacts, care must be exercised in collecting, reviewing, and judging student work artifacts.

The following concepts can aid in finding and using assessment artifacts.

1. Identify assessment artifacts among existing student work products and common behaviors.
2. Match the artifacts to particular educational goals.
3. Determine how the artifacts can be collected, e.g., representative samples of student work from required courses, unobtrusive but systematic observations of student behavior, alumni/ae records, etc.
4. Develop a tentative plan for small teams for reviewing and appraising the artifacts.
5. Determine if the discovered artifacts help the faculty and administration understand the actual outcomes of the MDiv program in terms of the students' overall growth.
6. If this assessment strategy adds value by increasing understanding, continue to rework it so that it can be sustained in terms of time, energy, and funds.
7. Identify the educational goals for which no relevant artifacts were found. At this point, decide if the value of special, add-on tests and surveys warrants their cost in terms of time, energy, and funds.

Figure 7 provides a mechanism for relating various kinds of artifacts that may provide evidence related to student learning in the four primary areas of education in the MDiv program.

Figure 7
Areas of Learning and Assessment Artifacts

Artifact Types	Religious Heritage	Cultural Context	Personal and Spiritual Formation	Ministry and Public Leadership
Papers				
Examinations				
Projects				
Intern Reports				
Spiritual Journals				
Psychological Assessments				
Survey Summaries				
Student Tracking*				
Other				

* Some seminaries, particularly Roman Catholic schools, maintain extensive, detailed records of students' intellectual, spiritual, personal, and leadership growth that are reviewed annually or semiannually by either the entire faculty and administration or a designated subset. These records could provide most, if not all, of the information for these seminaries' assessment programs.

Using this matrix, faculty members and administrators can identify potential artifacts, and a smaller group could review the artifacts to identify those that appear to be most relevant for each area. This smaller group could review samples of each selected artifact to determine each one's relevance to the particular area each is supposed to reflect. Once this group has selected a set of artifacts, a jury—possibly including outside readers—could read all or a sample of the artifacts. The jury's goal is to reach consensus about what a particular set of artifacts suggests about students' general progress in religious heritage, cultural context, personal and spiritual formation, or ministerial and public leadership.

A very good approach is to compare similar artifacts of first year students to second and/or third year students. If, for example, in religious heritage, a faculty desires information about students'

ability to think theologically, it might identify major assignments related to theological thinking in required courses in the different years. These assignments or artifacts could be read to determine the general level of theological reasoning they portray. Ideally, a faculty might track the same students longitudinally from their beginning year until graduation. Obviously, this process will require three years to complete. But when time and circumstance do not permit longitudinal tracking, a faculty could compare first, second, and third classes at the same time.

Principles of Assessment

As in any field, there are key principles of assessment that should guide and shape any assessment plan. One will recognize the similarity of these assessment principles and the general principles of scholarship and research in other academic fields.

Validity. The question of validity in assessment is: Does this test, survey, or artifact reflect the knowledge, values, understanding, or disposition being evaluated? There are several ways to estimate the validity of an artifact or assessment instrument.

1. Is there a logical connection between a given assessment instrument and the quality being assessed? Given the instrument and a description of the quality to be assessed, would informed and thoughtful individuals see a logical connection between the two? If asked to judge a test on principles of worship, for example, would independent observers agree that the test is faithful to the relevant theological and liturgical principles of the faith community?

2. Is there an empirical relationship between the assessment instrument and the quality it is intended to assess? Empirical validation may be gauged by how well a performance on an assessment correlates with other assessments of the same quality, ability, or attitude. For example, one could compare a student's

performance on a question-and-answer test on principles of worship to an exercise in which the student submits an order of worship reflecting these principles.

3. Does performance on an assessment instrument predict performance in the field? Using the same example of a test on principles of worship, do those that score the highest on the test also demonstrate that they achieve better ratings in conducting worship in their intern or field experiences?

The main point to consider in determining the validity of assessment efforts is that the evidence provided by the assessment instruments should accurately reflect the skill, knowledge, or ability that is being assessed. Three general ways of determining validity have been suggested: (1) In the judgment of informed and credible reviewers, does the assessment reflect the quality being assessed? (2) Do test scores, survey responses, and ratings of artifacts correlate with other measures of the quality being assessed? (3) Do assessment indicators predict later performance? A valid assessment indicator reasonably represents the knowledge, understanding, skill, trait, value, or behavior it assesses.

The purpose for this general, non-technical review of validity in assessment is also a cautionary reminder that assessment instruments and techniques do not necessarily assess what they are purported to measure. Elaborate psychometric analyses are not needed for all locally constructed tests, surveys, or protocols, but some effort should be made to evaluate the validity of any instrument or technique used consistently for important feedback to students or judgments about them. Some attempt should be made to estimate validity in at least one of the ways noted above.

Suggestion 4. Review and verify the validity of any assessment strategy, technique, or instrument consistently used to evaluate the quality of student learning and formation resulting from the MDiv curriculum or program.

Reliability. A test or survey that yields widely varying results when given to the same individuals several times is not reliable or consistent. An unreliable test is no better at assessing student knowledge or skill than an inaccurate rifle would be as a measure of marksmanship. If the same students respond quite differently to a survey of values on a first and second administration within a week, when no effort was made to change their values, it would make more sense to assume the survey is unreliable than to conclude students' values change over short periods of time. It is especially important to determine the reliability of surveys, tests, or other judgments of student work when the results are used to evaluate programs. Using an inconsistent or unreliable assessment technique or instrument to assess program outcomes is like measuring a room for carpet with an elastic tape measure.

Consistency or reliability is an issue in essay tests widely used in theological education. Persons who are experienced in grading these kinds of tests have become aware that they tend to mark the first papers they grade differently from the last ones they grade. A way to minimize this inconsistency in grading is to grade all the responses to one question, then all the responses to another, varying the order in which the tests are read with each question. It may also help with consistency of grading to make a list of the key points expected in a complete response to each item. Studies of teacher-constructed and teacher-marked essays do not demonstrate strong reliability when subjected to empirical tests of reliability.

This kind of inconsistency can also be a problem with assessing the degree to which an MDiv program is achieving its intended outcomes. The more a faculty understands about reliability or unreliability of locally constructed examinations and single-person grading, the less it will base indicators of program outcomes on fine distinctions. When reviewing artifacts from courses or major comprehensive projects to determine if students are generally meeting MDiv goals, it is important to remember that judgments

using broader, more general categories are usually more trustworthy than judgments based on finely distinguished categories.

Reliability or consistency increases with multiple readers making judgments. Studies have found that multiple readers' marks tend to vary less after they have marked several papers and worked together. They can learn to grade more consistently by discussing their different marks. Examinations and papers presented for degrees in the United Kingdom have for many years been marked by two or more readers. Multiple readers may mark papers without being explicit about their individual approaches or developing common rules among themselves, but reliability is enhanced when they become more conscious of the reasons they have for marks they give. In any case, multiple readers will need and want to find ways to increase their inter-rater consistency.

Certain marking practices tend to develop more inter-rater reliability than others. In general, multiple readers tend to agree more when the following practices are adopted:

- Make fewer distinctions. For example, if asked to sort papers into three stacks, such as Outstanding, Acceptable, and Unacceptable, there will be greater agreement than if graders are asked to sort them into five or more stacks.
- Holistic marking. Rather than elemental marking in which many rubrics are developed for content and composition, graders will be more consistent if they sort papers on holistic judgments. On the other hand, elemental marking is critical when the purpose of the marking is to give students feedback for improvement.
- Ranking rather than rating. Multiple readers will be more consistent in ranking student work from best to worst than in rating them on a five-point scale.

Standardized examinations are rarely used in theological education. However, if they have been properly developed, they are

likely more reliable than professor-designed tests, particularly if they are published by reputable test companies. Information on the reliability and validity of most published tests can be found in the *Mental Measurements Yearbook (MMY)* published by the Buros Institute of Mental Measurements. Interpreting the evaluations in the MMY requires some knowledge of measurement statistics. As mentioned earlier, the assessment instruments for which reliability is the greatest concern are personality tests, values inventories, and opinion surveys.

A technocratic culture is accustomed to assigning numbers to human qualities in general and to student performance in particular. It is also easy to place too much trust in these numbers. In *Technopoly*, Postman observes: "Our psychologists, sociologists, and educators find it quite impossible to do their work without numbers. They believe without numbers they cannot acquire or express authentic knowledge."[16] As in the case of the validity of an assessment, it is important to be mindful that assessment strategies may be inaccurate or unreliable.

Suggestion 5. Review and verify the accuracy or reliability of any assessment strategy, technique, or instrument consistently used to evaluate the quality of student learning and formation resulting from the MDiv curriculum or program.

Qualitative and Quantitative Information

Good assessment takes a "both/and" not an "either/or" perspective with regard to qualitative and quantitative information. Where numbers can reliably and validly contribute to understanding some aspect of seminary education they should be used. Larger realities should not be reduced to meaningless bits of information for the sake of quantitative assessment. Anecdotal information and subjective observations that have been gathered systematically and carefully and subjected to second- or third-party verification may provide meaningful (valid) and accurate (reliable) assessment.

Quantitative information is often derived from objective tests or surveys. An assessment instrument is considered objective if different readers score the responses alike; that is, the test or survey yields the same score regardless of who marks it. The most objective tests or surveys allow respondents to select only certain answers or options (such as multiple choice and true/false tests and surveys). They do not include items where the respondent composes an answer or response. There are many different ways such tests can be constructed and formatted. While objective tests often measure lower levels of knowledge and thinking effectively, developing sophisticated objective measures of deeper levels of understanding usually requires more training and skill in measurement than most faculty members have.

The more students supply answers to open-ended, essay questions, the less objective the test becomes because different readers will mark them differently. Consistency of final marking is improved by having multiple readers reach consensus on each paper's grade, but this consumes considerable faculty time and energy. Multiple-reader marking is usually reserved for high stakes, comprehensive examinations, projects, and theses. The level of agreement among faculty members when they mark the same papers indicates the extent to which they have become a faculty community, not just a collection of individual scholars.

As qualitative research has increased in social science, qualitative assessment now plays a significant role in higher education evaluation. Qualitative assessment may particularly interest theological educators because theology not only deals with the ineffable but also with inner attitudes and dispositions that influence behavior patterns. In addition to surveys, qualitative assessment strategies such as ethnography, unobtrusive observations, interviews, and individual and group narratives can be very helpful in assessing what *Michael Quinn Patton discusses as* "reflexivity":

Reflexivity has entered the qualitative lexicon as a way of emphasizing the importance of self-awareness, political/cultural consciousness, and ownership of one's perspective. . . . Being reflexive involves self-questioning and self-understanding. . . . To be reflexive. . . is to understand an ongoing examination of *what I know* and *how I know it*. . . . Reflexivity reminds the qualitative inquirer to be attentive to and conscious of the cultural, political, social, linguistic, and ideological origins of one's own perspective and voices of those one interviews and those to whom one reports.[17]

Many theological schools have cultural anthropologists or sociologists of religion who could provide helpful professional expertise in the use of ethnographic or similar strategies for assessing educational goals. There have been two ethnographic studies of theological education published in the past twenty years; and one of them, *Being There: Culture and Formation in Two Theological Schools*,[18] used its ethnographic investigation to explore the educational process and its apparent outcomes. The other is *Equals Before God: Seminarians as Humanistic Professionals*.[19]

Summary observations about quantitative and qualitative assessments follow:

- Objective tests and surveys are more reliable or consistent than essay, open-ended-item tests or surveys.
- Objective tests and surveys tend to measure lower-level thinking and responses.
- Objective tests and surveys can involve deeper levels of thinking, understanding, dispositions, or values, but developing these kinds of objective tests and surveys requires psychometric knowledge, experience, and skill.
- Qualitative assessments can more readily deal with higher levels of reasoning and deeper levels of feeling, but because they are subjectively marked, they are less consistent or reliable and usually fewer subjects are surveyed.

- The lower reliability or accuracy of qualitative assessments can be countered by systematic observations, multiple readers, and rubrics to guide graders.

Suggestion 6. Use both quantitative and qualitative assessment strategies and instruments.

Ability and Behavior

Theological schools tend to focus on abilities, such as theological knowledge and reasoning; facility with biblical languages, interpretation, and explanation; oral and written communication; and understanding of history and cultures. To assess these abilities, theological educators construct tasks or tests by which students can demonstrate knowledge, understanding, and skills. Because grades depend on students' performance on these tasks and tests, they typically put forth their best efforts on them.

Few people, however, always do their best. For example, ministers may not always use their best exegetical skills in preparing for weekly Bible studies. Typical behavior can be an important indicator of values, beliefs, and character. How a student consistently treats department store clerks may say a lot about how that student values others. In general, psychological inventories ask respondents to report typical, common, or recurring behaviors and thoughts. They do not serve their purpose if persons report what they think they should do or think.

Questionnaires and Surveys

Because questionnaires have such a major place in seminaries' assessment programs, they deserve special attention. ATS administers two questionnaires, the *Entering Student Questionnaire* and the *Graduating Student Questionnaire*, as well as a resource for assessing anticipated ministerial perceptions or behaviors, *Profiles of Ministry*. These instruments, others of similar quality, and student work deserve a major place in a seminary's assessment program.

Lee J. Cronbach distinguished between maximum and typical performance tests of ability.[20] Typical performance assessments can be divided into two types: behavioral observations and self-report devices. A student may be observed in ministry by persons in field work settings, by field supervisors, by peers, and by seminary faculty and staff. Self-report instruments, such as inventories or surveys, provide efficient and systematic means through which students can report their perceptions or understanding of values, feelings, or opinions. Survey results need to be interpreted with caution because of the pervasive tendency of respondents to respond positively. Self-report instruments are not the only ones with this positive bias. Many faculty want to give students the "benefit of the doubt" in grading their work; letters of recommendation often reflect a positive bias, as do personnel evaluations. Some faculty members worry about student ratings of their teaching that are in the mid-range—a 3 on a 5-point scale. Positive bias can lead to grade inflation and overly high ratings of faculty or institutional performance.

There are several ways to address the biases that tend to influence the results of surveys. A school can use published surveys or instruments, such as the ATS questionnaires, that provide average scores of students from many schools, which can serve as a point of reference for interpreting student ratings in one school. Writing survey items for maximum clarity, minimal bias, and respondent-friendly format requires some technical skill and considerable art. If there are no published surveys that address the issues a school wants to investigate, it might be able to contract with someone experienced in drafting, conducting, and analyzing surveys to help design the survey and advise on administration, analysis, and interpretation. Theological schools related to colleges or universities can likely find the requisite skills in faculty from business, education, psychology, political science, or sociology. There are many books and Internet resources on survey design and research; there are even tools on the Internet for construction, administration, and analysis of surveys.

As ATS schools seek resources to assess student learning, they will often turn to surveys, and as these instruments are given the power by the school to indicate whether a major learning goal has been attained or not, it is important that the school both use survey instruments that minimize some of the problems with surveys and gain skill in their interpretation.

In addition to respondent fluctuation and positive bias, survey assessment can have problems because the persons completing the survey do not represent the whole group or universe of persons from whom the results are desired. There are two ways to address this potential problem. One is to require everyone in a targeted group (all entering or graduating students) to complete the survey. Many schools do exactly this on surveys of populations over which they can exercise this kind of control. The second way to address the problem, when the population to be surveyed cannot be required to complete it (like graduates five or ten years out of school), is to draw a sample from the desired population. More valid conclusions about a population can be drawn from an 80 percent response rate of participants drawn into a sample of a population than a 20 to 30 percent response from the entire population. In general, it is better to have a high percentage of a small but representative sample than to get a larger number of non-representative responses. A smaller sample usually makes it easier to follow up on non-respondents and, eventually, get a response from them. Smaller samples also make telephone interviews more feasible.

Inevitably, most assessment programs will include surveys, but it is questionable to use them as the *only* source of feedback. Surveys and other ways of collecting information on the usual or typical behavior offer the greatest value when used as guides for conversation. In the case of comparing the reported values of first- and third-year students, the results could be summarized and formatted in ways that facilitate reflective discussion on the greatest changes as well as those values on which students report little or

no change. Having survey results before beginning a reflective discussion can add value in several ways:

- Students or other stakeholders may identify issues that they are reluctant to mention directly or individually.
- Survey data provide information that conversation alone may not, because the opinions of the most verbal participants can dominate a group discussion.
- Representative survey results can reveal the breadth of affirmation or concern about an issue.

There are no sure-fire ways to optimize discussions of surveys when they bring unexpected or bad news. Nevertheless, some approaches can lead to more useful discussions than others.

- Having an outsider with experience in analyzing and presenting survey results organize and present them.
- Having the same individual or another outsider facilitate discussions using group exercises that allow or actually encourage the timid as well as the verbally aggressive to contribute.
- Arranging for dialogues or "debates" in which certain individuals are asked to argue for understandings that they actually oppose.
- Concentrating on finding root causes before proposing solutions to problems.
- Not playing the blame game.

Suggestion 7. Use survey results as a checklist of issues worthy of reflective discussion to arrive at a deeper understanding on which to construct and implement sound changes.

5.   Assessing the Results of Information Regarding Learning Outcomes

As suggested earlier, the MDiv degree should be evaluated prima-
rily by its learning and formation results. The purpose of collecting
quantitative and qualitative assessment results is to prompt and
guide faculty, administrative, and stakeholder evaluative reflec-
tion on the value of the resources or institutional attributes and
how they are deployed in the educational program. Figure 8
illustrates the use of assessment results to reflect on resources and
program.

Figure 8
Assessment of Results for Evaluation of Program and its Resources
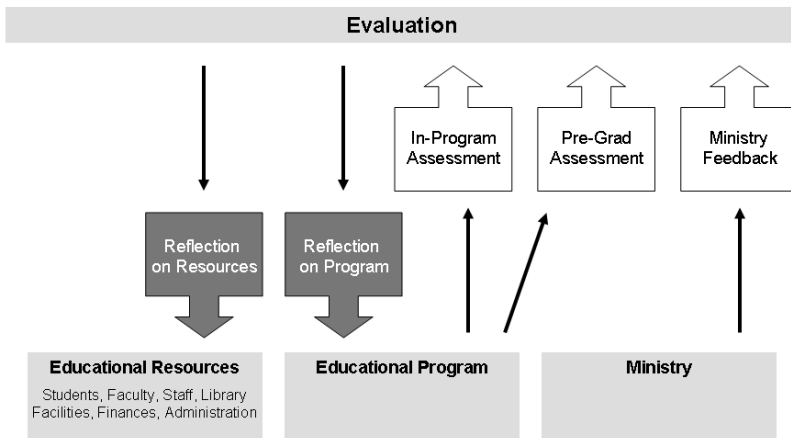
Evaluation by Corporate Reflection

Individual reflection on the findings of an assessment effort is less
than common and, given academic individualism, corporate re-
flection can be all but impossible. The very mention of another
committee meeting, whose agenda is "corporate reflection" on the
results of evaluating learning outcomes, is not likely to be greeted
with enthusiasm. Nevertheless, corporate reflection on informa-
tion about what students have learned is necessary because data

alone cannot place value on anything. Assessment is necessary; it is the process by which human beings, individually and corporately, reflect on data and information about learning in order to interpret it in the context of the purposes of a degree program. A faculty may corporately reflect on the results of efforts to evaluate student learning in several ways.

Compare Outcomes to Goals. Perhaps the most common reflective evaluation compares observed to desired outcomes. Are the students reaching the education and formation goals the program is designed to facilitate? In addition to the collective reflection of the resident faculty, respected outsiders might be asked to compare actual with hoped-for outcomes.

Compare Current and Past Results. As suggested earlier, the only real value of assessing results is to see if programs are achieving intended results and to use discrepancies between actual and intended results to stimulate and guide change. In almost all cases, change should not be based on one-time assessments. Major changes should be based on patterns or trends that are observed over time, not on limited or single observations. For example, if a seminary decided that one of its MDiv goals is that students will be able to preach effectively, as measured in field education contexts, changes shouldn't be made to the instructional strategies based on the evidence presented by one graduating class. One year's data cannot indicate a systematic trend. Program changes should be based on recurring patterns, and it is not clear whether the results of one year's evaluation are an exception or a common trend.

Benchmark Outcomes. Benchmarking involves one seminary comparing its assessment results to those of other seminaries that they identify as peer institutions, and it can provide a helpful basis for corporate reflection. Benchmarking is often used in two ways: it can mean comparing measurable outcomes in one school with measurable outcomes in another, comparable school, or it can mean comparing processes or procedures. In the first case, for

example, one school might compare the level of exegetical skill of its students after completing the required Greek courses with the skill of other students after comparable courses in other seminaries. In the second case, one school might compare its strategies and courses for teaching Greek with the strategies used by other schools. The practice of benchmarking is an effective way to use assessment for improvement. Increasingly, educational institutions are using benchmark strategies for both processes and results. Benchmarking is most effective when a school has a complete and detailed understanding or "map" of its own process. For example, if a faculty committee wished to benchmark an instructional program against another one, it should first understand the details of its own program. Graphic, detailed maps of instructional programs or administrative processes make it more possible to identify gaps and redundancies that might be otherwise overlooked. The process of mapping—or developing a flowchart of a program or process—also helps the school understand the process itself.

Distinguish Exceptional from Typical Outcomes. Academic institutions tend to cite their exceptional achievers as marks of quality. For example, an undergraduate college may cite two Rhodes scholars among its 2,000 graduates over the last ten years as evidence of its academic strength. However it is the usual student, not the exceptional one, who more accurately reflects the results of a program or system. For example, a monumental study of all Pennsylvania college students' general knowledge and intelligence was conducted from 1928 to 1932. A sub-analysis of engineering students' scores revealed an interesting and pertinent anomaly: "… the highest-scoring student, an ambitious youth of foreign birth, should hold the top score for the state in the general-culture test but should ascribe his knowledge not to local opportunities but to a constant and diligent study of the Sunday edition of *The New York Times*."[21] In this case, the student believed his extraordinary test performance resulted from extracurricular reading. An institution or program is more appropriately judged by its average

rather than extraordinary results. This principle is consistent with established statistical principles and practices in tracking outcomes in manufacturing, health care, and all institutions where the quality of a system is tracked in terms of its results. When tracking performance indicators, it is important to concentrate on patterns over time, not unusual or exceptional changes. Assume that all of a seminary's graduating MDiv students have taken a comprehensive theological examination from 1996 through 2005, with the results as shown in Figure 9. The average of the yearly averages from 1996 to 2005 was 78 percent, but the average score in 2005 was 60 percent. There could be many reasons for an unusually low score for one year that have nothing do with the overall effectiveness of the program. As Aristotle in *Nichomachean Ethics* wrote, "One summer does not a summer make." Changing entire systems should be based on established patterns of poor performance not one bad case.

Figure 9
Comprehensive Exam Scores
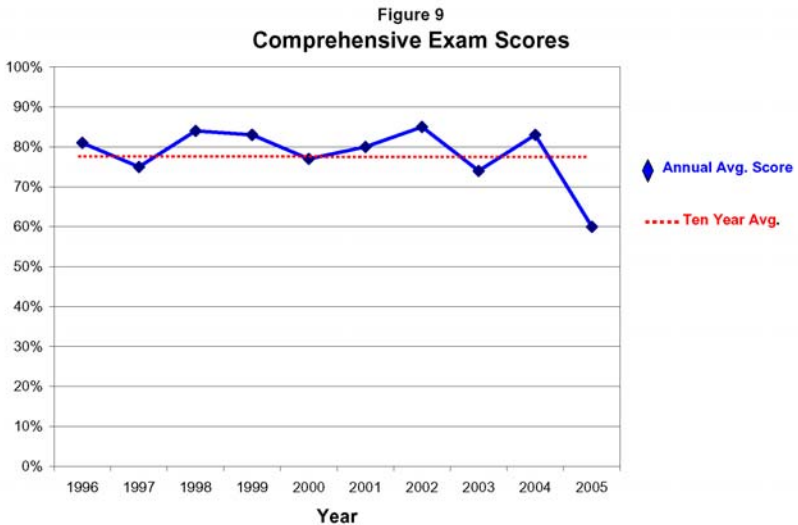
Search for Root Causes. However a faculty chooses to reflect on the findings of assessment information, the central focus of the reflection should be on discerning root causes of evidence of strength, success, weakness, or problems. Educational changes should be made on the basis of the best estimate of root causes. Assessment results rarely, if ever, reveal root causes of problems; at best, they alert interpreters to problems. They are like the part of the iceberg above the water; there is more below the surface than above it. Too often, educators jump from indications of problems to quick fixes or solutions based on our particular biases, rather than searching for root causes, or pursuing the deeper texture of theological learning.

One way to search for root causes is for a group of faculty to engage in a disciplined cause-and-effect pattern of brainstorming, as illustrated in Figure 10. This process helps a group brainstorm about possible causes before leaping to conclusions. As team members suggest causes, the facilitator writes a brief summary of each suggestion near its most appropriate category. Debating a suggested cause before the brainstorming is complete tends to dampen participation. Discussion of the suggested causes should begin only after the group has identified all reasonable suggestions. Next, persons should be encouraged to ask for clarification, to lobby for particular suggestions, and to question suggested causes. When this discussion has run its course, participants should engage in some decision-making process that leads to reasonable conclusions about the most likely actual causes. Then a team can begin its investigation with the top two or three.

Figure 10
Cause-and-Effect Diagram
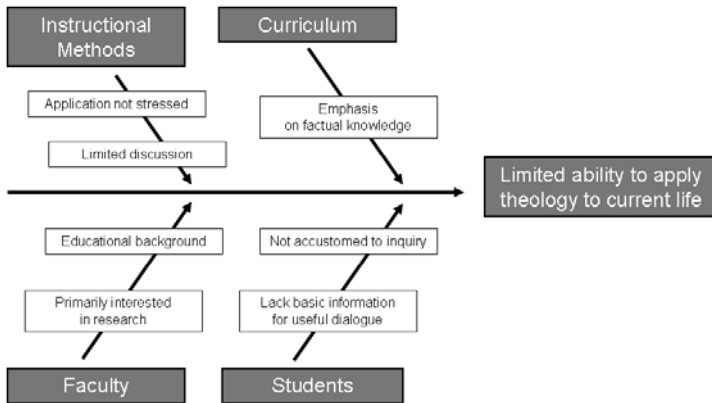
Note: The notes connected with each of the four legs of the diagram are terse, reminder statements of more complex thoughts. For example, the note "Emphasis on factual knowledge" might represent, "Emphasis on factual knowledge to the exclusion of the integrated learning expected in the MDiv curriculum."

The basic value of techniques such as the Cause-and-Effect Diagram helps a group focus first on causes before proposing a solution.

Suggestion 8. When corporately reflecting on assessment data and information use effective group processes and focus on root causes before proposing solutions.

6.   Revising Goals or Activities Based on the Assessment

If, as suggested earlier, assessment adds value when it ensures quality and prompts and guides improvement, that is where it most often fails. Yet assessment is often done to satisfy agencies outside the seminary, such as ATS. For this reason, accrediting organizations often require evidence that assessment has led to actual changes.

41

Change Goals in Light of Outcomes

This section begins with considering an unlikely strategy: aligning goals with results. Most typically, educators assume that they should try to improve some institutional or educational process to increase the chance that observed results more closely reflect goals. The fourth step in the assessment model described in the ATS standards suggests that the school should evaluate both its "activities" and its "goals." It could be possible that, after assessing outcomes with reasonable validity and reliability over time and finding that the goals are never met, the school might conclude that its goals are unrealistic. For example, suppose a school has the goal that all MDiv students have an advanced knowledge of church history across the entire Christian era, but the crowded curriculum allows for only one required course in church history. After repeated indicators that students do not graduate with an advanced knowledge of two millennia of church history, the school might conclude that the goal is too high; that is, it is simply not possible for students to attain the desired learning goal, given the limited space in the curriculum for church history courses. This frequent result might lead an institution to consider revising stated goals after observing a consistent pattern of results failing to meet expectations.

Change Activities to Reach Goals

Most often, goals are not changed. More typically, the results of assessing learning outcomes identify where goals may have not been met. After some careful assessment of probable root cause for the findings, the school may consider appropriate improvements in its educational strategies. As simplistic and obvious as it may appear, the Plan, Do, Check, Act Cycle is "a custom more honored in the breach than the observance." Faculty might be reluctant to make a significant change given past experience. In many cases, a change is never changed. For example, a past change could have been a quick fix that caused more problems than it solved. But once in place, it remained unchanged; its effectiveness and efficiency

were not continuously evaluated. In some cases, faculty and staff could be more open to change if convinced from experience that every change has to prove itself. The simple process of planning, doing, checking, and then acting again can ensure stakeholders that changes will be checked on and modified as needed. Peter Senge and his associates have identified pitfalls to avoid in making changes. Those proposing a change might consider a checklist of possible pitfalls when developing an improvement.[22]

1. Is the change a quick fix that will likely backfire?
2. Will the change cause growth beyond capacity to sustain it?
3. Will the change shift the burden to another unit?
4. Will the change draw upon common, limited resources?
5. Will the change create adversaries?

Incremental and Innovative Change

Change may occur in increments and through innovation. Incremental change modifies a program or system in small steps. For example, a lecture mode of instruction may be incrementally changed by adding PowerPoint, smart boards, or electronic query and response systems. In contrast, innovation refers to change on a large scale, like shifting a primary teaching strategy from lecture to problem-based learning. Assessment, as typically practiced in higher education, more often drives incremental change than innovation. This is not to say that assessment cannot drive paradigmatic change, but organizations rarely make dramatic changes just because results indicate they should. Innovation—large scale change—is more often driven by crisis. Furthermore, assessment strategies are usually designed to identify needed improvements, not collect information that questions the program itself. As Thomas Kuhn pointed out in *The Structure of Scientific Revolutions*,[23] physical scientists often do not see data that disagree with their paradigms. Joel Barker in *Discovering the Future: The Business of Paradigms*, applies Kuhn's conclusions to the difficulty organizations have in recognizing and dealing with paradigmatic shifts.

Academics are not immune to what Barker calls "paradigm paralysis," which he labels as "a terminal disease of certainty."[24] Barker explains the difficulty of innovating from inside as follows:

> Why is it that internal innovation is so difficult? Because *the* paradigm is already in place. So, until we change that attitude and stimulate people to be more flexible and break out of the paradigm paralysis to search for alternatives, we will continue to find the great new ideas, on the whole, being discovered outside the prevailing institutions.[25]

Organisms and Change

Increasingly, organizational theorists and practitioners view organizations as organisms rather than machines. Thomas Petzinger explains this shift as follows:

> As a model for everything, Newtonianism, it turned out, had a problem: It worked only within the narrow range of Newton's instruments. The "laws of nature" fell to pieces in space, as Einstein's relativity showed, and at the subatomic level, as quantum physics showed. Scientists realized that however useful in solving smooth, mechanical problems, Newton's calculus was meaningless in understanding the vast preponderance of nature: the motion of currents, the growth of plants, the rise and fall of civilizations. (Mechanics could explain why the apple fell, as the physicist Per Bak once quipped, but not why the apple existed, much less why Newton was thinking about it.)[26]

Richard Pascale and associates[27] understand organizations as organisms coping with their environment. Organisms change when they experience disequilibrium with their environment. Because organisms or agents within them reorganize to deal with disequilibrium, attempts to design or redesign an organization mechanistically are not necessary and ineffective. Leaders will do better to

allow or induce disequilibrium that stimulates the agents in the organization to work together to make the needed changes. As disequilibrium stimulates change and change improves the odds for survival, Pascale and associates believe equilibrium leads to stasis which brings eventual demise. Assessment adds little value if tacked on to traditional organizational structures. An organization understood and operated as a horizontal, flow-through system, as illustrated earlier in Figure 5, Optimizing Outcomes, can more readily use assessment for improvement than an organization of loosely coupled silos. Yet horizontal systems tend to be designed and managed as "machines." "Organize to Optimize" describes how an organic approach to higher education could optimize student learning.[28] By permission of the Observatory on Borderless Education, this paper is on the ATS Web site at www.ats.edu.

In conclusion, assessment's only value is in the improvement it stimulates and guides. Assessment makes its greatest contribution when an MDiv program is understood and organized as an integrated system focused on optimizing student learning.

Suggestion 9. Begin to develop a corporate understanding of the MDiv program as a horizontal, organic system that learns about its results to improve.

Conclusion: Assessment of Learning and Commission Accreditation

Generally, higher education institutions have ensured quality by documenting that they have met "industry" standards, e.g., faculty credentials, library collections, and financial integrity and sufficiency. In contrast, institutions are now expected to provide evidence that student learning outcomes approximate program goals. This guide is based on the assumption that MDiv programs are expected to provide evidence that graduates as a group have

attained the stated learning goals. In addition to ensuring quality in terms of reaching goals, seminaries are also expected to use assessment to prompt ongoing improvement. Most accrediting agencies, including the Commission on Accrediting of The Association of Theological Schools, expect direct evidence of quality and improvement through assessment of outcomes. Figure 11 contrasts documentation of quality assurance and quality improvement by attributes and outcomes.

Figure 11
Quality Assurance and Quality Improvement
by Attributes and Outcomes

	Attributes	Outcomes
Quality Assurance	The following attributes meet or exceed commonly accepted standards: • faculty credentials • library holdings • ratio endowment to students	Assessed learning outcomes approximate learning goals; in the case of significant disparities, there are plans for their elimination or reduction.
Quality Improvement	Over time, there are gains in • faculty credentials • library collections • endowments • facilities • faculty publications	Over time, there are steady gains in student learning outcomes.

While institutional attributes support learning, they do not necessarily ensure it. Similarly, enriching traditional attributes does not necessarily cause improvements or gains in student learning outcomes. Consequently, accrediting associations now expect institutions and programs to provide evidence that goals related to student learning have been attained. Improvement is gauged more by attainment in student learning than by enrichment of resources, i.e., faculty scholarship, library holdings, or faculty-student ratios.

Quality Assurance Expectations

Because theological schools are required to provide evidence that students in general reach stated learning outcomes, the assessment of the MDiv degree program in an institutional self-study could be evaluated by its treatment of the following five aspects of assessment:

1. Statements of discernible student learning goals in each of the four MDiv areas—Religious Heritage, Cultural Context, Personal and Spiritual Formation, and Capacity for Ministerial and Public Leadership
2. Assessments of outcomes of graduates' learning in each of the four MDiv curricular areas during the program, before graduation, and following graduation
3. Systematic comparison of each stated learning goal with its respective average or typical assessed outcome
4. Identification and description of each significant disparity between a stated, expected learning goal and its respective assessed outcome
5. An explicit, feasible plan to reduce each significant disparity

Visiting teams are asked to review seminaries' assessment programs against the five criteria suggested above, and they might use questions such as the following to evaluate assessment efforts:

1. Student Learning Goals
    a. Does the seminary have explicit learning goals for each of the four ATS MDiv areas: Religious Heritage, Cultural Context, Personal and Spiritual Formation, and Capacity for Ministerial and Public Leadership?
    b. Do the learning goals reflect the emphases of the faith community or communities served by the seminary?
    c. Are the learning goals logically related to the seminary's mission?
    d. Are learning outcome goals stated so that seminary faculty and independent, external reviewers can discriminate with reasonable consistency between student work that meets or fails to meet them?

2. Assessment
    a. Does the seminary provide evidence of direct measures of the achievement of student learning outcomes rather than relying on indirect measures such as satisfaction surveys

or other indices such as strength of library holdings, faculty credentials, and published scholarship?

b. Are there assessment strategies for each goal?

c. In comparing each assessment with its student learning goal, is there an apparent, logical connection between the goal and its assessment?

d. Other than apparent, logical correspondence between goals and their respective assessments, is there evidence of the validity of various assessment tools or strategies?

e. When a sub-group is used to obtain assessment data, what evidence exists that the subgroup represents the group as a whole? *This issue is particularly important for survey data because in most cases only a modest percentage of those surveyed complete and return the surveys.*

f. When ratings of student work are used as assessment strategies, does the school have strategies in place to compare ratings and assess the consistency of raters' markings?

g. Have marking guidelines or rubrics been developed and shared with readers?

h. When student papers, projects, and student work artifacts are reviewed by two or more readers, what is the evidence that the readers rate student work products similarly or consistently?

i. When a sample of student artifacts or portfolios is reviewed to gauge achievement in particular learning outcomes, what evidence is offered that the selected artifacts and the process for reviewing them validly assess the outcome(s) in question?

3. Has each learning goal been systematically compared with its respective, average or typical, assessed outcome?

4. Descriptions of Comparisons

a. Have the congruence or disparity of assessed outcomes with respective student learning goals been accurately described?

    b.   Do the self-study descriptions appear to be consistent with the visiting team's understanding of the student learning goals and the assessments they have reviewed?

    c.   How has the seminary demonstrated that it is dealing thoughtfully with disparities between student learning goals and actual outcomes?

    d.   How do the school's priorities of concern correspond to the best judgment of the visiting team?

5.   Improvement Plans

    a.   Does the seminary explicate a plan to reduce each significant disparity between learning goals and assessed outcomes?

    b.   Does the plan address the root cause of the disparity?

    c.   Given the seminary's mission, constituency, and resources, does the plan appear to be appropriate and feasible?

    d.   Does the seminary provide opportunities for reflection upon the results of its assessment strategies and does it document the results that emerge from these reflection opportunities for the improvement of student learning?

ENDNOTES

1.   Victor Klimoski, "Assessment and Good Teaching," *Teaching Theology and Religion*, 8, no. 2, 2005: 69-77. Carolyn Jurkowitz, "What is the Literature Saying about Learning and Assessment in Higher Education?" *Theological Education* 39, no. 1, 2003: 53-93. Victor Klimoski, Kevin J. O'Neil, C.Ss.R., and Katarina M. Schuth, O.S.F., Editors and Respondents, *Educating Leaders for Ministry: Issues and Responses* (Liturgical Press, 2005). Daniel Aleshire, "The Character and Assessment of Learning for Religious Vocation: MDiv Education and Numbering the Levites," *Theological Education* 39, no. 1, 2003: 1-15.

2.   "Potemkin Village" comes from an apocryphal story that Grigory Potemkin built fake villages in an attempt to impress Empress Catherine II of Russia with his development work in the Ukraine as she rode through the countryside. "'Potemkin village' came to denote any pretentious facade designed to cover up a shabby or undesirable condition." "Potemkin, Grigory Aleksandrovich, Prince Tavrichesky, Imperial Prince." *Encyclopædia Britannica*, 2005, Encyclopædia Britannica Online, August 4, 2005. http://search.eb.com.ezproxy.samford.edu/eb/article-9061055.

3.   John Harris and Dennis Sansom, *Discerning Is More Than Counting*, AALE Occasional Papers in Liberal Education #3 (Washington, DC: The American Academy for Liberal Education, March 2000). Available online at http://www.aale.org/pdf/harris-sansom.pdf. Hard copies may be obtained from The American Academy for Liberal Education, 1050 17th Street NW, Suite 400, Washington, DC 20036.

4.   Michael Polanyi, *Personal Knowledge: Towards a Post-Critical Philosophy* (Chicago, IL: The University of Chicago Press, 1962), 2.

5.   Michael Power, "The Perils of the Audit Society," George Jones and Steve John, eds. (LSE Public Service Seminar Series: Government Department, London School of Economics and Political Science, 1997), 8.

6.   David Boyle, *The Tyranny of Numbers: Why Counting Can't Make Us Happy* (New York: HarperCollins Publisher, 2001) and *The Sum of Our Discontent* (New York; London: Texere, 2001).

7.   Harris and Sansom, op. cit., 4.

8.   Master of Divinity Degree Program Standard A, A.2 Primary Goals of the Program and A.3.1 Content, *Bulletin 46*, Part 1, 2005 (Pittsburgh: The Association of Theological Schools), 193.

9.   Lee S. Shulman, "Making Differences: A Table of Learning," *Change* 34, no. 6 (Washington, DC: Heldref Publications, November/December 2002), 31. Reprinted with permission of the Helen Dwight Reid Educational Foundation. Copyright 2002.

10.   Dennis Sansom, *Being Christian in Ethics: How Distinctive Features of Christianity Shape Ethics* (unpublished manuscript), Department of Philosophy, Samford University, 2005.

11.   Oxford English Dictionary, 2nd Edition, online.

12.   Ibid.

13.   Elliot W. Eisner, "Instructional and Expressive Educational Objectives: Their Formulation and Use in Curriculum," 1967, ERIC #: ED028838.

14.   Ibid. 8, 15, 16.

15.   Neil Postman, *Technopoly: The Surrender of Culture to Technology* (New York: Vintage Books, 1992), 13.

16.   Ibid.

17.   Michael Quinn Patton, *Qualitative Research & Evaluation Methods,* 3rd ed. (Thousand Oaks, CA: Sage Publications, 2002), 64-65.

18.   Jackson Carroll, Barbara Wheeler, Daniel Aleshire, and Penny Long Marler, *Being There: Culture and Formation in Two Theological Schools* (New York: Oxford University Press, 1997).

19.   Sherryl Kleinman, *Equals Before God: Seminarians as Humanistic Professionals* (Chicago, IL: University of Chicago Press, 1984).

20.   Lee J. Cronbach, *Essentials of Psychological Testing,* 3rd ed. (New York: Harper & Row, 1969), 29-35.

21.   William S. Learned and Ben D. Wood, *The Student and His Knowledge: A Report to the Carnegie Foundation on the Results of the High School and College Examinations of 1928, 1930, and 1932* (New York: The Carnegie Foundation for the Advancement of Teaching, 1938), 251.

Appendix

Six Assessment Books Selected from
2004 Assessment Institute's
Book Descriptions
October 31 - November 2, 2004
Indiana University Purdue University in Indianapolis

Assessing Student Competence in Accredited Disciplines:
Pioneering Approaches to Assessment in Higher Education
Catherine A. Palomba and Trudy W. Banta
$37.00, ISBN: 1579220347 – www.styluspub.com

How do accreditors encourage attention to the assessment of student learning? How do faculty on the campuses respond to these actions by accreditors? What can we learn from the accredited disciplines about defining learning outcomes, promoting thoughtful campus planning for assessment, and improving student competence?

Specialized accrediting bodies stand at the forefront of the assessment movement in higher education and exert important influences on program and institutional assessment and improvement. The academic programs these bodies approve are frequently among the first on a campus to consider ways to assess student competence. This book focuses on their approaches to assessment.

The book opens with two foundational chapters. The first examines the concept of student competence and reviews the historical and political contexts in which assessment takes place. The second reviews best practices in carrying out assessment, including the important roles of faculty development and assessment leadership.

Case studies then describe how faculty in eight professionally oriented disciplines have developed and practiced assessment on

their campuses and the lessons these faculty offer to colleagues in their own and other disciplines. The book concludes with chapters on the use of authentic assessment within several disciplines on a single university campus, the British quality assurance movement, and a review of the lessons, issues, and challenges that are common across disciplines.

The stories and examples presented in this book—covering issues as diverse as faculty and student involvement, the role of outside stakeholders, and balancing external guidance and institutional autonomy—will provide immediate practical guidance for faculty and administrators active in assessment regardless of their discipline.

Assessing Student Learning: A Common Sense Guide
Linda Suskie with Foreword by Thomas A. Angelo
$40.00, ISBN: 1882982711 – www.ankerpub.com

Interest in assessing student learning at institutions of higher education—and the need to learn how to do it—skyrocketed in the last two decades of the twentieth century and continues to grow into the twenty-first century. This book summarizes current thinking on the practice of assessing student learning in a comprehensive, accessible, and useful fashion. Short on background and theory and long on practical advice, this is a plainspoken, informally written book designed to provide sensible guidance for assessment practitioners on virtually all aspects of student assessment, and for faculty who simply want to improve assessments within their classes. *Assessing Student Learning* presents readers with well-informed principles and options that they can select and adapt to their own circumstances.

Assessment Clear and Simple: A Practical Guide for Institutions, Departments, and General Education

Barbara E. Walvoord

$25.00 - ISBN: 0-7879-7311-4 – www.josseybass.com

Assessment Clear and Simple is "Assessment 101" in a book—a concise, step-by-step guide written for everyone who participates in the assessment process. This practical guide helps to make assessment simple, cost-efficient, and useful to the institution, while at the same time meeting the requirements of accreditation agencies, legislatures, review boards, and others. The book explores a variety of topics and shows how to:

- Build on assessment already in place
- Use classroom work and grading process
- Get faculty and departments on board
- Assess hard to define goals such as moral and civic development
- Development of workable learning goals
- Tailor assessment to its purposes
- Select sensible assessment measures
- Make criteria explicit
- Use assessment to improve leaning
- Establish effective oversight without an assessment bureaucracy
- Write an assessment report
- Interpret the institution's culture to external audiences

Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education

Catherine A. Palomba and Trudy W. Banta

$38.00, ISBN: 0787941808 – www.josseybass.com

This step-by-step guide provides the most current practices for developing assessment programs on college and university campuses. *Assessment Essentials* outlines the assessment process from the first to the last step and is filled with illustrative examples to

show how assessment is accomplished on today's academic campuses. It is especially useful for faculty members and others who may be new to the assessment process. In clear, accessible language, Palomba and Banta describe effective assessment programs and offer a thorough review of the most up-to-date practices in the field.

Assessment in Practice: Putting Principles to Work on College Campuses
Trudy W. Banta, et al.
   $45.00, ISBN: 07879013420 – www.josseybass.com

   Assessment in Practice brings together in one volume the best current knowledge of which assessment methods work best and how their principles should be incorporated into all effective assessment efforts. Whether at institutional, program, or departmental levels, drawing from 165 actual cases, and reporting eighty-six of them in their entirety, the authors illustrate methods and techniques of assessment covering a wide range of objectives in diverse types of institutions.

Building a Scholarship of Assessment
Trudy W. Banta, et al.
   $35.00, ISBN: 0787959456 – www.josseybass.com

   In this book, leading experts in the field examine the current state of assessment practice and scholarship, explore what the future holds for assessment, and offer guidance to help educators meet these new challenges. The contributors root assessment squarely in several related disciplines to provide an overview of assessment practice and scholarship that will prove useful to both the seasoned educator and those new to assessment practice. Ultimately, *Building a Scholarship of Assessment* will help convince skeptics, who still believe outcomes assessment is a fad and will soon fade away, that this is an interdisciplinary area with deep roots and an exciting future.