

SELF-STUDY HANDBOOK

## CHAPTER SEVEN

*A Reflective Guide to Effective Assessment  
of Student Learning*



The Association of Theological Schools  
The Commission on Accrediting

## TABLE OF CONTENTS

<b>Preface</b>	<b>1</b>
<b>Introduction</b>	<b>2</b>
<b>Section 1: The Purpose and Importance of Assessment</b>	<b>4</b>
<b>Section 2: Some Limits and Challenges to Assessment</b>	<b>7</b>
<b>Section 3: A Definition and Description of Assessment</b>	<b>11</b>
<b>Section 4: Three Guiding Questions for Effective Assessment</b>	<b>14</b>
a. Is it simple?	14
b. Is it sustainable?	17
c. Is it sufficient?	19
<b>Section 5: Four Foundations of Effective Assessment</b>	<b>25</b>
a. Identification of desired goals or outcomes	25
b. Information on the achievement of those goals or outcomes	28
Quantitative and qualitative information	29
Direct and indirect measures	30
Benchmarks or performance indicators	34
Rubrics	35
Demography and delivery	36
c. Interpretation of that information by key players	37
d. Implementation of those interpretations for meaningful improvement	39
<b>Section 6: A Checklist for Effective Assessment</b>	<b>43</b>
<b>Appendix 1: Sample Summaries of Assessment Results</b>	<b>45</b>
<b>Appendix 2: Excursus on Quantitative Data</b>	<b>48</b>
<b>Appendix 3: Excursus on Rubrics</b>	<b>50</b>

# Preface

The title of this document, “A Reflective Guide to Effective Assessment of Student Learning” is intentional for these reasons:

- As a **reflective** guide, it is more than just a fact-based set of instructions for doing assessment. Rather, it contains practical and personal reflections regarding perceptions and practices of assessment. As such, it is less formal than other sections of the *Handbook*, using, for example, the editorial “we” throughout. The intention of this more personal approach is to make assessment more understandable to a wide range of constituents, including administrators, staff, faculty, and trustees.
- As a reflective **guide**, it suggests one way to do effective assessment but by no means serves as “the guide” or “the only way” to do assessment. The ATS Commission Standards of Accreditation (“Standards”), not this guide, constitute the final word on assessment for member schools. Still, schools that review this guide, based on those Standards, should be well positioned to meet the expectations on assessment of the ATS Board of Commissioners (“Board”). That Board, which approved this guide, is the representative body of peers and public members elected by the Commission membership to make accrediting decisions on behalf of the entire membership.
- Addressing **effective assessment**, this guide does not (and cannot) encapsulate every ideal of assessment. It is based on best practices of assessment recognized by the ATS Board of Commissioners to be effective and is founded on the Commission Standards on assessment that frame and inform those best practices.
- Regarding assessment **of student learning**, this guide focuses almost exclusively on how member schools assess what students learn, rather than on the broader area of institutional evaluation, although the two are clearly related and, on occasion, interrelated, as we discuss in section 3.

# Introduction

**Assessment.** It is a simple educational idea, to reflect thoughtfully on the work of teaching and learning. Yet assessment can also strike fear into the hearts of deans, trigger resentment from faculty, and throw self-study coordinators into confusion and despair. Perhaps that is why assessment is, by far, the most frequent area of needed growth (and concern) surfaced by accreditation committees and by the ATS Board of Commissioners—all composed of your peers. It seems to be an area of struggle across many graduate institutions of theological education—regardless of size, faculty, finances, denominational affiliation, longevity, or any other factor. We believe, however, that assessment can be one of the most rewarding tasks that a school undertakes. This little guide is intended to highlight the high value that assessment can have for your school. Please note that in this guide *assessment* refers primarily to *assessment of student learning*, as opposed to the broader but related arena of institutional evaluation (see section 3).

While we assume that all ATS member institutions have some experience in assessment, we also know that many individuals involved in assessment are often new to their institutions or new to the role of assessment, given the normal transitions among member schools. These transitions involve those who serve not only as assessment directors or coordinators but also as presidents and deans—and faculty and trustees. We mention the latter two especially because the ATS Commission Standards approved by the membership in 2012 make it clear that “the buck stops” for assessment in two places: with the faculty and with the board. Educational Standard, section ES.6, Assessment of Student Learning Outcomes, includes two statements that reinforce this point: (1) “Evaluation of student learning is the responsibility of the faculty” (ES.6.4.1), and (2) “The governing board of the school is responsible for ensuring that the school has a program of assessment of student learning and that the results of these activities inform and shape educational and institutional decisions” (ES.6.4.3).

This guide attempts to demystify assessment in at least two ways. First, it provides a concise and practical framework to help institutions imagine, implement, and improve their own assessment programs. Second, it serves member schools by clearly linking the project of assessment to the expectations of the ATS Commission Standards.

As we begin, an important caveat is warranted. The ATS Board of Commissioners, peers elected by the entire ATS Commission membership, believes there is *no single best way* to do assessment, but there are *best practices* for effective assessment that are highlighted by the Standards and discussed in this document. This is called “A Reflective Guide to Effective Assessment of Student Learning” because it

*This guide provides a concise and practical framework to help institutions imagine, implement, and improve their own assessment program, and it serves member schools by clearly linking the project of assessment to the expectations of the ATS Commission Standards.*

is intended as just that. It is a guide to assessment deemed *effective*—in light of the ATS Commission Standards, approved by the membership of approximately 250 Accredited Member or Candidate

*We offer this brief guide not to break new ground in assessment but to highlight some best practices (based on the Standards of Accreditation) that we see among member schools—and to warn against some not-so-good practices that we sometimes see.*

for Accredited Member schools.

The literature on assessment is extensive, even exhausting. We offer this brief guide to highlight some best practices (based on the Standards) that we see among member schools—and to warn against some not-so-good practices that we sometimes see. This is not

a quick fix for schools that struggle with assessment, but it is meant to be a helpful guide along the assessment path, especially the path that ATS member schools walk as they focus on assessment of student learning within the context of graduate theological education. Like other sections of the *Self-Study Handbook*, this document seeks to be illustrative and suggestive. It is intended to help schools think about how to implement the Commission Standards, specifically those dealing with the assessment of student learning. In the end, every school's assessment efforts will be evaluated in light of the Commission Standards. We believe, however, that this guide will help schools better understand and implement the Standards regarding the assessment of student learning.

## Section 1

# The Purpose and Importance of Assessment

Assessment and its twin, evaluation, appear nearly 300 times throughout the Standards, beginning with General Institutional Standard 1, culminating in Educational Standard, section ES.6, and emphasized in the first section of all ten Degree Program Standards. Why all this concern about assessment? [For a discussion of the terms *assessment* and *evaluation*, see section 3.]

A key reason why assessment of student learning merits more attention and raises more concerns, by far, than any other issue before the ATS Board of Commissioners is this: assessment helps schools know how they can better achieve their mission. Nothing matters more than mission. That is why virtually every accrediting agency lists mission or purpose first among their Standards, including the ATS Commission Standards. We even heard of one hospital administrator whose title on the door read “Vice President of Mission Achievement.” She was their director of assessment.

*Assessment is not about compliance but about a common commitment to continuous improvement in our educational practices and our institutional missions. Assessment is about improvement.*

The primary purpose of the ATS Commission on Accrediting is to “contribute to the enhancement and improvement of theological education through the accreditation of schools that are members” (Commission on Accrediting *Bylaws*, Section 1.2). It is the nearly 250 institutional members of the Commission that have developed and approved the Standards—and their increased focus on assessment. The Standards are intended to “identify qualities associated with good institutional and educational practice” among member schools and to “articulate the shared understandings and accrued wisdom of the theological school community” (*Preface to the Standards of Accreditation*). It is crucial to understand at the outset that *assessment is not about compliance but about a common commitment to continuous improvement* in our educational practices and our institutional missions. Assessment is about improvement.

To be sure, one oft-cited reason to care about assessment is that it is required to be accredited. And accreditation—at least in the United States—is the gateway to more than \$150 billion in annual



federal financial aid, without which most US schools could not survive, including many ATS member schools. And in Canada, there is increasing provincial pressure to document quality assurance through assessment of student learning. However, to say that schools should do assessment because accreditation—or worse, the government—requires it is similar to faculty telling students to “learn this because it will be on the test.” If assessment is simply “on the test” (or “is the test”), some schools may fail or barely pass because there is not appropriate motivation to do the work (lack of assessment data) or to do it adequately (weak assessment data; e.g., only perception-based). Some may struggle simply because they do not understand the concept. Recall the Ethiopian official’s response to Philip in Acts 8:31, “How can I understand unless someone should guide me?” Assessment is not important because it will be “on the test” for accreditation. It is important because it helps you ask yourself these three fundamental questions:

- What are you trying to do (in light of your mission)?
- How do you (and others) know how well you are doing it?
- How might you do it more effectively?

Perhaps a more appealing way to think of assessment is this: Assessment is about telling your institution’s story. Or better, how to improve your institution’s story. And the heart of your story is your mission. Solid assessment data help you know and tell your story, especially when you can blend numbers and narratives. Which assertion in each of these pairs of statements conveys a more powerful story?

- Our students make great pastors; or,  
Over the last five years 95 percent of our MDiv graduates have been placed in more than 50 different ministries around the world.
- Our students like our school; or,  
Our students rate their experiences here in the top tier of all seminaries across North America.
- We’ve got a great faculty; or,  
Our students consistently cite our faculty as the number one reason why they came—and why they stay.

*Assessment is simply asking yourself what it is that you are trying to do and then thoughtfully considering how well you are doing it and how you might do it better.*

Lest we be accused of pandering to mere marketing, we understand that assessment is about improving student learning, not “bragging” about what you do well. However, part of assessment is holding yourself accountable to share assessment information—good and bad—with key constituents in appropriate contexts. Some stories will celebrate your strengths. Other stories may be candid confessions of concerns that need addressing: We are not yet where we want to be, but here is where we are going—to help our students improve and better achieve our mission.

Effective assessment tells your *unique* story, instead of a generic or idealized story, because it asks the questions *you* care about, based on criteria that are defined by *you* and that matter to *you*. Good assessment draws on what *you* already know and what *you* value. It is about beginning where you are, not reinventing the wheel or fitting into someone else’s box. Assessment is certainly not about bowing to bureaucracy or being deluged by “administrivia” (as one dean once described it). Assessment is part of the natural curiosity of educators and institutions. It is simply asking yourself what it is that you are trying to do and then thoughtfully considering how well you are doing it and how you might do it better.

The focus of assessment, then, is about improvement and accountability—to yourself, your school, your students, and your constituents—not simply to an external entity. It really is a matter of institutional integrity: are you doing what your mission claims? *Assessment demonstrates to your stakeholders—including students and staff, donors and denominations—how well your school is living out its mission.* Assessment also helps institutions address the critical issues of where to focus limited resources, where to expand programming or personnel, and how to plan for the future.



## Section 2

## Some Limits and Challenges to Assessment

*Assessment is a terrific servant but a terrible master.*

As important as assessment is, it is only fair to discuss its limits and its challenges. The first limit we would list is that assessment is not a panacea for solving all of a school's problems. We believe it is a very important part of what a school does, but it is not everything that a school should do nor even the most important thing. Some ATS member schools, for example, must focus so much energy on addressing their financial problems that assessment of student learning cannot be their main priority. As one accreditor on an evaluation visit to a troubled school commented, "You can't focus on the effectiveness of your furnace when your house is on fire." And yet, even here we would note that assessment is a key factor in addressing financial issues effectively, as it provides data that can help an institution determine which expenses might be cut, which programs to expand, and which budget priorities to address. Put another way, assessment itself can't put out the fire, but it can sometimes help a school identify which fire-suppression strategies to attempt.

Related to this, we often hear from schools that one of the key challenges in assessment is that it can become an all-consuming, even overwhelming task. While it is a key emphasis in the ATS Commission Standards, it is not the only focus. To be sure, assessment is important, but it is meant to be a means to a greater end, not an end in itself. Faculty cannot spend so much time on *assessing* what students learn that those energies diminish *what* students learn. One colleague with rural roots put it so succinctly: "You can't fatten a hog by weighing it." Assessment is a terrific servant but a terrible master.

Another limit to assessment is often echoed in this familiar statement: *Not everything that is valued can be measured, and not everything that is measured should be valued.* While assessment typically requires some means of measurement (usually stated in terms of both multiple measurements and multiple measurers), the ATS Commission Standards revised in 2012 focus more on "demonstrability" than "measurement." Assessment cannot be reduced to or limited by some mechanistic, formulaic set of numbers. Measurement that evokes an image of a

*Grades can certainly be a part of a school's assessment efforts, but accreditors and the public—your peers and your constituents—long ago stopped viewing course grades as sufficient to tell the whole story of what students are learning.*

fixed yardstick or simple scale is invariably inadequate to assess the complex concepts foundational to theological education. Many of the most important things that ATS member schools care about seem to defy measurement, such as forming students personally and spiritually. However, that does not mean they cannot be discerned in some way and demonstrated to key constituents—internally and externally—from donors to accreditors. Just because something is hard to assess does not mean it is impossible to assess or not worth trying. To be sure, it sometimes requires some form of proxy. For example, one cannot truly know whether students are growing internally in their walk with God, but one can observe certain behaviors that are indicative of such growth within one's particular context, such as time spent in private devotions, participation in public worship, involvement in ministry activities, interactions with peers and professors, and so on.

Sometimes schools will begin an assessment plan by thinking of what it is they can easily count or track, rather than starting with the question of what it is they want to know and then working backward to find data that will help answer those specific questions. Unfortunately, this sometimes leads to too much data and not enough useful information. It also tends to minimize the complexity of assessing theological education. We are reminded of the man who lost his keys one night and was looking for them under a street light. A passerby asked him if he could remember the last place he had them, to which he responded, "Down the street." Asked why he didn't look for them there, the man replied, "The light is better here." Assessment needs to look even where the light is not good.

Yet another limit and challenge in assessment is "faculty resistance." While that may be dwindling, it is still real—especially in these challenging economic times when everyone seems to be doing more with less. Faculty, already feeling like the ancient Hebrews in Egypt ("more bricks, less straw"), may feel like this is just one more brick (or the last straw). Assessment is sometimes perceived by professors as another "unfunded mandate" from some accreditation bureaucrats and a "faddish" one at that. But if assessment is a fad, it has long legs—by North American higher education standards—going back at least 40 years. And given the present climate in Washington and in more and more provincial capitals in Canada, calls for increased quality

*Part of the joy and reward of effective assessment is having meaningful faculty conversations about how well students are learning what faculty feel they need to learn—what is working and what needs attention.*

assurance and accountability in documenting student learning will likely only get louder.

Other faculty argue that they have been doing assessment all along—in the form of grading. To be sure, *course grades are a form of assessment*, and they should not be simply dismissed as of no value. They can certainly be a part of a school's assessment efforts, but accreditors and the public—your peers and your constituents—long ago stopped viewing course grades as sufficient to tell the whole story of what students are learning. Four reasons for that come to mind: (1) course grades can be rather arbitrary (just ask any student), (2) course grades usually signify the perception of an individual instructor rather than the evaluative consensus of faculty as a whole, (3) course grades are notorious for inflation (“Grades here run the gamut from A- to A+”), and (4) course grades focus on individual students and individual courses, rather than on the goals embedded in an entire degree program. Course grades may tell part of the story, but much is lost (including cocurricular experiences and the progress and cumulative effect of a student's learning) if these receive too much weight. Similarly, tracking all course (or even assignment) grades can be distracting or can lead to more data than might be useful.

Related to the course grade challenge is the tendency of some faculty to focus their assessment efforts on individual students, rather than to see assessment also as a faculty-wide enterprise that looks at student learning institutionally and not just individually. The danger for many is to see only the trees and not the forest. This challenge will be discussed later in the first part of section 5.

Part of the joy and reward of effective assessment is having meaningful faculty conversations about how well students are learning what faculty feel they need to learn—what is working and what needs attention. As one faculty member finally acknowledged: “Assessment isn't *extra* work; it is our work.” It is part of the vocation of faculty. As teachers, we want to know if our students are learning what we hope they are learning. As scholars, we are trained to evaluate texts and other works to see where they are strong and where they are weak. Assessment draws on the faculty's natural curiosity (Did my students learn this?), educational passion (I want my students to learn this!), and scholarly expertise (I know what learning looks like).

*Assessment of theological education is no less a challenge and no less a reward than poetry—more art than science.*

Perhaps the most important limit and challenge to assessment is that schools will get so involved in doing it that they forget *why* they are doing it. Some of the more disheartening examples of that are assessment plans that on paper seem to have every “jot and tittle” of what the books say assessment should have but in principle and practice have no “heart and soul” of why the effort is even worth it. Such plans are all science and no art, written only to satisfy an accreditation requirement, with no chance of making any difference in those schools’ purposes or people or programs.

Such mechanistic approaches to assessment recall a scene early in the movie, *The Dead Poets Society*. A new teacher, played by Robin Williams, has a student read aloud the introduction to a textbook on poetry. The book’s author mechanically wrote that every poem could be plotted formulaically on a scale of “(1) perfection and (2) importance, thus determining its greatness.” Williams’ character, chagrined at the mechanical rigidity of this form of assessment, tells his students to rip that page out of their textbooks with this memorable line: “We’re not laying pipe. We’re talking about poetry. I mean how can you describe poetry like American Bandstand? Well, I like Byron, I give him a 42, but I can’t dance to it. . . . [Poetry] reminds us that the powerful play goes on, and you may contribute a verse.” Assessment is not numbers on a Likert scale. Assessment of theological education is no less a challenge and no less a reward than poetry—more art than science. “We’re not laying pipe. We’re talking about . . . people and purpose and passion—about mission and ministry.” *That is why assessment is important. What will your verse be?*

## Section 3

# A Definition and Description of Assessment

*The General Institutional Standards focus on institutional evaluation, and the Educational Standard and Degree Program Standards focus on assessment of student learning. We mention both here briefly, but the subject of this guide is assessment of student learning, not institutional evaluation—though the two are clearly interrelated.*

In one of the clearest treatments on the topic, *Assessment: Clear and Simple*, Barbara Walvoord defines the term this way: “Assessment is the systematic collection of information about student learning, using the time, knowledge, expertise, and resources available, in order to inform decisions that affect student learning” (2004, p. 2). One ATS member school defines it this way: “Assessment is a sustained effort to help us better achieve our mission by establishing mission-related goals for our programs, collecting and evaluating evidence around those goals, and then using that evidence to improve—our school, our students, ourselves.” Another member frames the assessment of student learning this way. “Assessment is asking ourselves these important questions: What do we want our students to learn (know/be/do) in light of our mission? How do we know whether or not they are learning it? And how can we use that information to improve their learning?”

Other schools have other definitions. In the end, definition is not nearly as important as implementation. Perhaps you can identify with this ATS dean: “Our difficulty is that we think we have resolved this just by talking about it. We need to *do* something.” Two leading experts in assessment highlight that very point in their book, *Assessment Essentials*, observing that the important question is not how assessment is defined, but whether assessment is used (Palomba and Banta, 1999).

The ATS Commission Standards do not have a precise definition of assessment but focus instead on how assessment should be used, including a framework for doing that well. The Standards talk about two related terms—evaluation and assessment—to describe the larger concept. The *General Institutional Standards* focus on *institutional evaluation*, and the *Educational Standard and Degree Program Standards* focus on *assessment of student learning*. We mention both here briefly, but the subject of this guide is assessment of student learning, not institutional evaluation—though the two are clearly interrelated, as we will see in the following paragraphs.

It may be worth mentioning here that some of the latest literature on assessment increasingly distinguishes between *student outcomes* and *student learning outcomes* (see, for example, *Higher Education Outcomes Assessment for the Twenty-First Century*, 2013). Student outcomes tend to focus on such things as graduation and placement rates, or the cost and time to complete a degree, including issues of student debt. Those are issues that might be more institutional in nature. Some of those may fall under what the Commission Standards refer to as institutional evaluation, though they may also legitimately fall under what the Standards call degree program goals because they deal with educational effectiveness (see 5.a below). Student learning outcomes tend to focus on what students actually learn through their curricular and cocurricular experiences. While the two are not unrelated, it is the assessment of student *learning* outcomes, rather than student outcomes (or institutional evaluation) that is the special focus of this guide. That said, the components of effective institutional evaluation are still central to effective assessment.

*Institutional evaluation* is described in General Institutional Standard 1 (section 1.2.2) in this way:

Evaluation is a *critical element in support of integrity to institutional planning and mission fulfillment*. Evaluation is a process that includes (1) the identification of desired goals or outcomes for an educational program, or institutional service, or personnel performance; (2) a system of gathering quantitative or qualitative information related to the desired goals; (3) the assessment of the performance of the program, service, or person based on this information; and (4) the establishment of revised goals or activities based on the assessment. Institutions shall develop and implement ongoing evaluation procedures for institutional vitality and educational effectiveness (emphasis added).

That introductory statement forms the foundation for all institutional evaluation—and for all assessment of student learning. It identifies up front the four foundational elements of any good evaluation plan or effective assessment process: (1) *identification* of desired goals or outcomes, (2) *information* about how well those goals or outcomes are being achieved, (3) *interpretation* of that information by key players,

especially faculty, and (4) *implementation* of those interpretations for the purpose of meaningful improvement (see section 5).

*Assessment of student learning* is described in the ATS Commission's Educational Standard, section ES.6.1, in this way:

A school shall maintain *an ongoing process for assessing student learning outcomes and degree program goals*. An effective plan of assessment should be as simple and sustainable as possible while adequate to answer fundamental questions about educational effectiveness. This plan should include (1) *a process for evaluating* components of the full degree program in an ongoing manner; (2) the identification of appropriate *direct and indirect indicators* of student learning; (3) the routine *involvement of faculty* in the review and evaluation of the results of the assessment; and (4) *linking assessment results to curriculum and educational planning, institutional strategic planning, and resource allocation* (emphasis added).

We will spend the rest of this guide unpacking these two statements from the ATS Commission Standards. Before that, it may be helpful to say a brief word here about formative vs. summative assessment. **Formative assessment** is sometimes called assessment *for* student learning and focuses on *in-process* strategies that provide faculty immediate feedback on what is working and what is not (e.g., asking students in the middle of a class session or term or program if they “get” it). **Summative assessment** is sometimes called assessment *of* student learning and focuses on *end-product* strategies that provide faculty ultimate or penultimate feedback on whether students “got” what they were supposed to learn. Both are important, but this guide focuses on the latter.



## Section 4

# Three Guiding Questions for Effective Assessment

Educational Standard, section ES.6.1, quoted at the end of the previous section states that “*assessment should be as simple and sustainable as possible while adequate [sufficient] to answer fundamental questions about educational effectiveness.*” This highlights three key questions to guide effective assessment:

### 4.a Is it simple?

The ATS Board of Commissioners gives no extra credit for extra length. In fact, some of the *least* effective assessment plans it has seen are some of the *longest*. Given that half of all ATS member schools have fewer than 150 students, 10 full-time faculty, and three degree programs, it is just not reasonable nor necessary for most schools to have elaborate assessment plans and results that run into dozens, scores, or even hundreds of pages. Even for larger schools, it is important that clarity trump length. Some of the most effective assessment plans that the ATS Commissioners have seen are only a few pages long, with assessment results summarized in one or two pages per degree program (though the raw data behind those summaries might be more extensive).

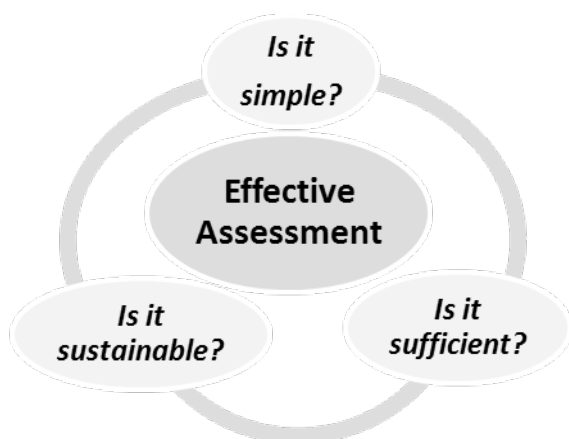
*If faculty are not seriously engaged in assessment with meaningful conversations about the results, then the plan is pointless and the results useless.*

In part, this is a pragmatic concern: faculty simply do not have the time to pore over pages upon pages (or screen after screen) of assessment data. And if faculty are not seriously engaged in assessment with meaningful conversations about the results, then the plan is pointless and the results useless. Such a plan and process would also not meet the ATS Commission’s Educational Standard, section ES.6.4, which “requires that faculty review and analyze the results of the assessment activities, discern appropriate changes to curriculum and educational practices, and document the results . . .” (section ES 6.4.1). To be sure, a school may *designate a person* as its assessment coordinator to help the faculty do assessment, but a school *may not delegate assessment to any one person* or office in order to bypass the faculty as a whole.

A key strategy to keep assessment simple is to keep the first factor in assessment simple, namely the number of degree program goals or

student learning outcomes (see the discussion on these two concepts in 5.a below). Schools need to *think carefully about how many goals or outcomes are appropriate for each degree*, particularly in light of their mission and context. The ATS Board of Commissioners has seen too many “best-laid plans . . . go awry” because schools tried to assess too

much. Many ATS member schools have found it useful to focus on four to seven goals or outcomes per degree program, not 14 to 17, even though each goal or outcome may have several components (see 5.a below). Schools that name too many goals (20 or 30 or more) often have a hard time adequately assessing those goals, and they may also struggle to be able to name or explain them to stakeholders.



Goals or outcomes are not meant to be a list of everything that is important to you or everything that you hope your students will learn. (Individual classes will still have their own course objectives

to cover more specific learning outcomes.) Rather, they should focus on key factors that give a picture of a successful graduate and help demonstrate that you are doing what you say you are doing, including attention to the program content areas described in the applicable Degree Program Standard. Sometimes the first step in assessment is to do less, not more. Cutting the number of degree program goals or student learning outcomes may be the first step in keeping your assessment efforts simple.

In addition to simplifying your list of goals or outcomes, another strategy toward a simple assessment plan is to remember that most schools already have a wealth of ongoing assessment information but may not realize it. Some of this information may exist within artifacts that are created in individual courses but not be known or reviewed by the faculty more broadly. For example, a senior reflection paper or an integrative paper in a class typically taken just prior to graduation or a sermon or case study at an internship site might provide very useful data about a particular degree program goal or student learning outcome. However, the assessment results for that assignment might be known only to the faculty member who assigns it. Documenting those already-reviewed results for broader discussion among the

*Sometimes the first step in assessment is to do less, not more. Cutting the number of degree program goals or student learning outcomes may be the first step in keeping your assessment efforts simple.*

whole faculty could be a good way to keep your assessment efforts simple.

Using these sorts of “course-embedded assessments” is an extension of the principle raised earlier not to reinvent the wheel, but “to begin where you are” by using what you already have. If you want to assess how well your MDiv students can exegete a first-century text for a twenty-first-century audience and you have existing MDiv course assignments that require students to do that, aggregate all of those student assignments for your overall assessment. You don’t need to develop an entirely new artifact or assessment tool. Other information might exist at an intuitive or occasional level but needs to be concretized and regularized within the life of the faculty. For example, faculty might discuss the curriculum when they are preparing to start a faculty search, or when something particularly good or bad or surprising happens, but not in the regular and ongoing way that good assessment planning requires.

Other data may exist but need to be synthesized and documented into an accessible, communicable format. For example, a new program would understandably not yet have any data on what graduates have learned. However, one can assume that a new program came about as a result of some form of assessment (faculty conversations, task force work, curricular discussions, student interviews, engagement with constituencies for whom the program is intended, etc.)—at least in terms of what the intended learning outcomes for the new program should be. All of those efforts represent valid assessment efforts. Those efforts may just need to be synthesized and documented for both internal and external reviewers. Schools are often criticized for not having a *culture of assessment*, when more often what is lacking is a *culture of documentation*. Such documentation allows a school to satisfy the Standards and the expectations of external reviewers, sets up a school to engage thoughtfully and intentionally in systematic review and analysis of data (it is hard to analyze intuition or institutional memory!), and also helps everyone to “tell your story well,” as mentioned earlier. Some examples of such documentation might include minutes or notes from faculty or board meetings, or proposals from curricular review committees. The first appendix to this document also provides more common examples of how schools might document assessment results.

*Schools are often criticized for not having a culture of assessment, when more often what is lacking is a culture of documentation.*

In terms of documenting assessment, the ATS Board of Commissioners does not require any particular assessment model, but many schools find the classic “Nichols model,” or some variation of his “5-column model,” to be a helpful way to maintain a *simple* assessment process. James O. Nichols, a university director of institutional effectiveness, pioneered in the early 1990s a five-step process for connecting the assessment of student learning outcomes to institutional effectiveness and to improvement. The full five columns include (1) institutional or departmental goals, (2) student learning outcomes, (3) means of assessment and criteria for success, (4) summary of assessment data collected, and (5) use of results. Nichols’ 1996 classic, *A Practitioner’s Handbook for Institutional Effectiveness and Student Outcomes Assessment Implementation*, has gone through several editions and iterations. Variations on the original Nichols model are widely used in higher education circles today (see Appendix 1 for examples). The key is to keep it simple, while still sustainable and sufficient/adequate—as we discuss below.

#### 4.b Is it sustainable?

To say that assessment should be simple does not mean it should be simplistic or easy. Good assessment takes careful thought and hard work—sustained over time. Again, as the ATS Commission Educational Standard, section ES.6.1, reminds us, assessment “should be as simple *and* sustainable as possible.” This leads to two key and interrelated questions: Is your assessment *plan* sustainable? And, are your assessment *practices* sustained over time? A key question for accreditors is not only “How are you doing?” but also “Where have you been and where are you going?” While a first essential step is for schools to have an assessment plan (see section 6), accreditors and other stakeholders long ago stopped being satisfied with simply a *plan* for doing assessment. It is assessment *practices and results* over time that matter, especially how schools use those results over time to make educational improvements for its future.

One of the more problematic (as well as least effective) assessment practices that evaluation committees sometimes discover is that a school has gone through an intensive assessment process to prepare for an evaluation visit, but only as a one-time event. Not only is there no schedule for ongoing review, but the amount of time and effort that went into the process is often not sustainable over time.

*Accreditors long ago stopped being satisfied with simply an assessment plan. It is assessment practices and results over time that matter.*

Unfortunately, at times such processes also create resentment rather than buy-in from stakeholders, such that doing it next time will be even harder, not easier. Just as with other practices (think *exercise*), it is more valuable to do regular and ongoing assessment than it is to do one intense attempt every now-and-again. In preparing assessment plans and strategies, then, a school will want to think about what makes sense in its institutional context, taking into account the workload and regular rhythms of its faculty and staff, among many other factors.

While the ATS Commission Standards of Accreditation are not explicit about how frequently each degree program goal or student learning outcome must be assessed, the *Educational Standard* and *Degree Program Standards* do use words like “ongoing” and “regular.” Schools will want to gather and assess data often enough that they can clearly demonstrate program effectiveness and can quickly assess the effect that curricular improvement or other changes (including faculty changes, course revisions, new textbooks, a shifting financial environment, and so on) are having on what students are learning. Perhaps not surprisingly, a best practice is to collect data every year and to have at least some of these data (one or more outcomes or goals) as the topic of substantial faculty conversation each year.

Incidentally, schools preparing self-study reports often ask how many years’ worth of data should be included in the report. While that varies due to several factors (e.g., time since last accreditation visit, nature of changes in the school, type of data being collected, etc.), there needs to be enough data to show what the trends are. And it typically takes at least three points (years) to plot a trend. Anything less doesn’t say anything about where you are going or where you have been, only where you are. As with other data presented in a self-study or ATS report, you will want to use (and analyze) enough data to show the path you’ve traveled to get where you are today, as well as how you’re making decisions to move forward from here. Remember again that the goal of assessment isn’t for assessment’s sake, but rather for ongoing educational and institutional improvement as you seek to live out your mission.

One last element of sustainability relates to the idea that evaluation of student learning is the responsibility of “the faculty” (ES.6.4)—as a whole, not just a committee or a few individuals. As noted earlier,

*Perhaps the most fundamental question to ask is whether your assessment helps your school achieve your mission.*

grading can be subjective; the same is true with assessment. In addition to identifying artifacts (or other sources for raw data), good assessment plans also include descriptive criteria so that assessment happens in a consistent fashion, even if personnel change. For example, rubrics (see 5.b and Appendix 3) can ensure that both students and faculty and any outside reviewers know clearly what is expected and what has been achieved. This also recalls the point raised earlier about course grades, namely that grades can be an effective means of assessing some things, but utilizing grading or scoring rubrics developed by several faculty is a better practice because it is more sustainable in terms of offering consistency and reliability. It also offers a broader lens through which to view what your institution means by degree program effectiveness.

#### **4.c Is it sufficient to answer fundamental questions about educational effectiveness?**

Perhaps the most fundamental question to ask is whether your assessment helps your school achieve your mission. That is why the ATS Board of Commissioners pays such close attention to member schools' assessment efforts. Assessment is about mission. For assessment to address the mission question, it requires that degree program goals or student learning outcomes are connected in some way to the school's mission.

Degree program goals, typically expressed in terms of student learning outcomes (but see discussion under 5.a), should reflect in some way the mission and ethos of that school. Doing so helps a school answer the most important question it can ask itself: Are we accomplishing our mission? That question also helps a school to align its assessment efforts with its missional context. Or, as Educational Standard, section ES.6.2.1, puts it: "The process and goals of assessment shall be conducted in ways that are congruent with the educational mission of the school."

We repeat here the point made earlier: effective assessment tells your story well because it asks the questions *you* care about, based on criteria that are defined by *you* and matter to *you*. Good assessment draws on what *you* already know and what *you* value—in the larger communal context of ATS Commission expectations set by all member schools. And what *you* value depends on *your* mission. For example,

a university divinity school with a strong missional focus on research (publish or perish) will likely assess its student learning outcomes differently than a Roman Catholic seminary whose primary mission is to prepare priests for ministry (publish or parish). The former might look at such things as the number of PhD graduates finding careers in the professorate or the number of student papers published, while the latter may consider such things as the number of MDiv graduates serving in parishes or the quality of homilies preached.

One strategy in linking assessment to mission is to develop a curricular map, though this approach may not appeal to everyone. Think of the classic periodic table of elements from chemistry class, as the sample table below illustrates. This strategy invites faculty to align graphically the core courses of a program with the goals or outcomes for that program. We suggest core courses because assessment focuses on all students' general experiences, whereas elective courses are often taken by only a few students. As illustrated in the table below, however, the curriculum is not necessarily limited to formal courses taught in a classroom (onsite, offsite, or online). Some parts of the curriculum—at least for ministry-oriented degree programs—are best delivered by experiences outside the classroom, such as supervised ministry or formational experiences. Those conversations typically require faculty to engage with administrators, particularly those who oversee those nonclassroom experiences. Those are still, however, important parts of a program's overall curriculum and need to be assessed and addressed.

Sample (and Simplified) Curricular Map					
	Course A	Course B	Course C	Experience D	How Often Goal Addressed
Goal/Outcome 1	X		X	X	3
Goal/Outcome 2	X				1
Goal/Outcome 3					0
Goal/Outcome 4	X			X	2
How Many Goals Addressed by Course	3	0	1	2	6



*The point of a curricular map is not for faculty to check boxes on a grid. The point is to get faculty talking about how well their courses help achieve their program's goals or outcomes, which, in turn, should help advance the school's mission.*

The point of a curricular map is to facilitate faculty conversations about how well the program's goals are covered by the program's curriculum. In the example above, faculty might want to discuss why Goal/Outcome 3 is not covered by any curricular requirements or why Course B does not address any program goal or outcome. It could be that Goal/Outcome 3 needs to be deleted, or some core courses need to be added or revised. And, it could be that Course B should no longer be core, or even be taught. Remember, however, that the number of boxes is "better weighed than counted," meaning that having only one box checked may be more than adequate, while having three boxes checked may reflect needless redundancy more than necessary reinforcement. To address such concerns, some faculty take the curricular map a step further and, instead of using simple "Xs" in boxes, use a more informative code, such as "I" (this course introduces that goal/outcome), "R" (this course reinforces that goal/outcome), or "C" (this course is the culmination of that goal/outcome in this program).

The point of this strategy is not for faculty to check boxes on a grid. The point is to get faculty talking about how well their courses help achieve their program's goals or outcomes, which, in turn, should help advance the school's mission. Remember that, while effective assessment often causes a school to revise its curriculum to achieve an outcome (per Educational Standard, section ES.6.4.1), sometimes assessment results can cause a school instead to change a goal or an outcome. For example, if the evidence suggests that students really are learning what faculty want them to learn in that program, even though what they are learning does not align with some current program goals or outcomes, then those goals/outcomes may need to be changed. The operative issue here is that any assessment plan needs to be not only simple and sustainable, but also sufficient: *adequate to answer fundamental questions about educational effectiveness.*

There are many ways to show that a plan is sufficient or adequate, most of which relate to some sort of demonstrable change, a closing of the loop. When the plan is not adequate, then it is the element that may need to change; as Educational Standard, section ES.6.2.4, notes: "Schools shall include in their assessment plans a comprehensive evaluation of the assessment plan itself. . . ." One of the least effective assessment reports that ATS Commissioners have seen was simply a

*The principle behind this Standard is that assessment is not only about improvement but also about accountability.*

submission of pages upon pages of raw data from student surveys—with no analysis (or even synthesis) of those results, no indication of what faculty did with those results, and no list of improvements made from those results. A school that discovers that it is collecting data that it does not analyze or analyzing data that never leads to change or improvement will likely find that it needs to assess and revise its assessment plan.

Before leaving this discussion about the third guiding question for effective assessment (namely, “Is it adequate [sufficient] to answer fundamental questions about educational effectiveness?”), it is worth noting the requirement described in the ATS Commission’s Educational Standard, section ES.6.4.4: “The institution shall, on a regular basis, *make available to the public a summary evaluation of the educational effectiveness of its approved degree programs*. The school shall determine the frequency and manner of this information” (emphasis added). Regarding “educational effectiveness,” this requirement raises the bar (and enlarges the audience) for assessment, since such information is to be made public. The principle behind this Standard is that assessment is not only about improvement but also about accountability—as we discussed in section 1 under the purpose and importance of assessment. That includes accountability to the various publics our member schools serve.

The ATS Board of Commissioners provides further guidance on this “public” requirement, as described in its *Policy Manual* (section V.G.3.c):

The Board understands Educational Standard, section ES.6.4.4 (regarding the institution making available to the public information regarding its educational effectiveness) to include *such data as time to completion, numbers of completers, and placement rates*], as well as *qualitative evaluation information indicating the educational effectiveness* of the school’s degree programs. Information regarding educational effectiveness may be provided in summary form as determined by the school (emphasis added).

Section 1.4.2 of each Degree Program Standard also requires “measures such as the percentage of students who complete the program and the percentage of graduates who find placement appropriate to their vocational intentions.” We understand that many schools do not

“place” students for a variety of denominational reasons. As part of the Annual Report Forms, the ATS Board of Commissioners, however, still requires member schools to track and report placement data for each degree program, regardless of who does the placing. For a discussion of the place of such things as placement rates and graduation rates in a guide on assessing student learning, please see 5.a in the next section.

It should be noted, however, that in asking for placement data, the ATS Commission on Accrediting does not assume that all degree programs have a vocational intent or that employment is always the best indicator of “student success.” We understand that many students enroll in member schools more for personal growth than for professional placement. In fact, the ATS Annual Report Form has a category intentionally called “non-vocational placement.” Nonetheless, placement data can still be an important indicator of educational effectiveness, with the understanding that placement does not simply mean “graduates got jobs.” Schools will want to connect placement rates to their own mission and context. For some schools, anything less than a 100 percent placement rate might be worrisome. For others, particularly for certain programs, placement might not be as useful (or as valued) an indicator regarding educational effectiveness.

It is also important to draw on other factors to fill in the story of placement. For example, are students perceived as being prepared or successful in their placements? Insights of those who supervise new graduates, as well as those served by new graduates, can be particularly useful here. Do students stay in their careers and excel or advance in the ways one might predict, or do they drop out or find only limited success in their work? By asking these sorts of questions, placement becomes a story, rather than just a statistic.

Many schools choose to make public in their statements of educational effectiveness not only direct measures expressed in quantitative terms (such as placement rates per degree program) but also indirect measures expressed in qualitative terms (such as synopses of student testimonials about their educational experiences or snippets of exit interviews). Some schools also provide results from student satisfaction surveys, which are typically indirect measures expressed in quantitative terms (e.g., our students rated their overall experiences here at 4.6 out of 5). Some combinations of all these can make a compelling

case, and help the school “tell its story well.” More information about direct and indirect measures and about quantitative and qualitative data will be presented in the second part of section 5.

## Section 5

# Four Foundations of Effective Assessment

The four foundations of effective assessment have already been introduced in section 3 above. They are found in General Institutional Standard 1, section 1.2.2, and may be summarized as follows:

1. *Identification* of desired goals or outcomes for each degree program
2. *Information* on the achievement of those goals or outcomes per program
3. *Interpretation* of that information by key players, especially by faculty
4. *Implementation* of those interpretations for meaningful improvement

### 5.a Identification of desired goals or outcomes for each degree program

The ATS Commission Standards frequently use two related terms: *degree program goals* and *student learning outcomes*. While these two terms are often used distinctively in the Standards, they are sometimes used synonymously. For example, each Degree Program Standard has separate sections for “Primary Goals of the Program” (section 1.2 of each) and for “Learning Outcomes” (section 1.3 of each). Section 1.3.1 of many of those Degree Program Standards explains the relationship between the two this way: “The primary goals of the program shall be further delineated as demonstrable learning outcomes.” On the other hand, their relationship can be so close as to be interchangeable, as evidenced by the very next line in section 1.3.1: “Institutions shall demonstrate that students have achieved the goals *or* learning outcomes of the degree program . . .” (emphasis added). While schools may choose to have separate goals and outcomes, perhaps the simplest way to think of these two terms is synonymously: student learning outcomes are a way to state degree program goals in terms that focus on what the school wants all students in that program to have learned by the time they graduate.

Regardless of the terms used, the assessment of student learning requires schools to use language that focuses on what students are learning (what it is that schools want students in a given program to know, do, be, feel, etc.). That typically requires active verbs that reflect some demonstrable behavior on the part of the students, such as “students will explain X” or “graduates will demonstrate Y” or “students will identify Z.” These brief examples are not meant to be exemplars of student learning outcomes. For clearer, more complete model outcomes statements, an Internet search on “student learning outcomes examples” will yield a wealth of help. Two of the more helpful sites are found at the [Center for Teaching and Learning](#) at Indiana University-Purdue University Indianapolis and at the [National Institute for Learning Outcomes Assessment](#) (NILOA).

At this point, it is worth pointing out possible exceptions to the principle just stated. There may be occasions when degree program goals are better stated in terms of the institution, rather than the student. For example, an institution may have as a goal that the degree program has a certain graduation or placement rate, or be financially sound or demonstrate diversity. Those are all appropriate goals for a degree program and appropriate issues to assess for educational effectiveness, but they would not be articulated as student learning outcomes. In addition, while student learning outcomes are to be communicated publicly “through the school’s catalog, website, and course syllabi” (ES.6.4.2), some institutionally oriented degree program goals are better kept as nonpublic, internal goals (e.g., financial results or admissions targets). To be sure, some would argue that such things as graduation or placement rates do not assess student learning at all but rather relate to institutional evaluation (see section 3 regarding *student outcomes* vs. *student learning outcomes*). Our point is that these can be “indicators of program effectiveness” (to quote the opening line to Educational Standard, section ES.6). As such, they are legitimate areas to review in determining how well a school is achieving its educational mission—a key purpose of this guide.

The previous point notwithstanding, student learning outcomes should be stated in terms of what schools want their students to learn (i.e., know, do, be, feel, etc.) as a result of completing the program. In doing so, it may be tempting to wax poetic or be overly aspirational in stating goals or outcomes (e.g., our students will change the world).

However, the *Degree Program Standards* require that degree program goals be “delineated as *demonstrable* learning outcomes congruent with the institution’s mission” (see, for example, sections A.1.3.1 or B.1.3.1). In writing a student learning outcome, it is important to use language that allows the school to demonstrate how well students in that program have achieved that outcome. For example, an outcome that “graduates of this program will be spiritually mature” could be hard to demonstrate, whereas “graduates of this program will demonstrate spiritual maturity in the following ways . . .” [and then list some ways that are important for that school’s mission and context] will allow a school to design assessment strategies focused on “those ways.”

*The most basic unit of assessment is the individual student, so that is a good place to begin—but not a good place to end. Effective assessment focuses on each and every degree program, not simply each and every student.*

It may be important at this point to highlight an important expectation in the ATS Commission Standards relative to assessing student learning. While student learning outcomes should focus on students, assessment of student learning is not primarily about *individual* students. We understand that it is important for a school to know if Jane or John or Juan is doing well as an individual student (not to mention how important that is to Jane or John or Juan). But we believe it is even more important for a school to know how well all of its Janes and Johns and Juans are achieving its degree program goals or outcomes. To be sure, the most basic unit of assessment is the individual student, so that is a good place to begin—but not a good place to end. Many ATS member schools excel at assessing individual students but fail to move from individual instances into institutional conversations about how well their degree programs are achieving their goals or outcomes for all students, and how, in turn, achieving those goals or outcomes help the school achieve its mission.

Before turning to the second foundation of effective assessment, it is important in identifying desired goals or outcomes to recall a point we raised earlier in section 4. To keep assessment simple (and sustainable), the number of degree program goals or student learning outcomes needs to be manageable. Many ATS member schools have found four to seven goals or outcomes per degree program to be a reasonable number, giving due attention to the program content areas described in section 2 of the *Degree Program Standards*. A word of caution may be in order here. To keep things simple, some schools may have two or more degree programs whose goals or outcomes overlap. While some



overlap is acceptable and even perhaps inevitable (e.g., three of the four content areas for the MDiv and the professional MA overlap), it is not acceptable simply to use the same exact set of degree program goals or student learning outcomes for more than one degree—with no distinctions in those degrees in terms of goals or outcomes. This point is highlighted in Educational Standard, section ES.1.1.1:

When Commission institutions offer more than one degree program, *they shall articulate the distinctions among the degrees with regard to their educational and vocational intent.* Institutions shall *articulate the goals and objectives of each degree program* they offer and assure that the design of its curriculum is in accordance with the institutional purpose and the Commission Standards of Accreditation (emphasis added).

In addition, while the number of goals or outcomes should be fairly “simple,” that doesn’t mean each goal or outcome cannot address complex issues. Some goals or outcomes may have several parts or “indicators” (see 5.b below). Nor does simplicity require schools to set the bar so low in wording a goal or outcome that achievement of that goal or outcome is virtually automatic for every student in that degree program. The bar is best set where the faculty (and key constituents) believe it should be set, even if that means some initial results may be a bit discouraging. To quote Browning, “Ah, but a man’s reach should exceed his grasp, Or what’s a heaven for?” On the other hand, the bar should not be set so high that degree program goals or student learning outcomes defy achievement. We are reminded of a doctoral advisor who discouraged anyone from writing a dissertation in her area of expertise with these disheartening words, “Only the impossible is worth doing, and you’re not qualified to do that.” In the following paragraphs we will discuss further how the goals or outcomes of a degree program can be evaluated and achieved.

*Collecting the right information is a critically important component of assessment.*

### **5.b Information on the achievement of those goals or outcomes per program**

It is worth noting that the *Educational Standard* has a separate section just on data (section ES.6.3). Collecting the right information is a critically important component of assessment. We recall one semi-nary dean who complained that his school all too often made important decisions on the basis of *the bias of the vivid example*. When his

colleagues met around the decision-making table, the “best story” often carried the day, even if that story was not at all normative or verifiable beyond that singularly vivid example. Collecting the right kinds of information is crucial, though that does not diminish in any way the power of a really good story—as long as meaningful data exist to support the memorable story. In the end, the information or data collected—whether a statistic or a story or both—must be sufficient to demonstrate how well the goal or outcome being assessed has been accomplished.

The section on assessment data (ES.6.3) highlights two important pairs of data or kinds of information, both of which are vital to understanding how well goals/outcomes are being achieved. Those two pairs are qualitative and quantitative kinds of information (ES.6.3.1), and direct and indirect measures (ES.6.3.2). This part of section 5 will look at those two pairs of information, along with three other important factors in collecting information, using these subheadings:

- Qualitative and quantitative information
- Direct and indirect measures
- Benchmarks or performance indicators
- Rubrics
- Demography and delivery

Regarding the overall collection of assessment information, this guide is too brief to discuss the myriad ways that schools might manage the data they collect. Those ways range from simple spreadsheets to proprietary programs on assessment that manage thousands of pieces of information. Here, too, the watchwords should be simple, sustainable, and sufficient.

### **Qualitative and quantitative information**

Both qualitative and quantitative information provide important and useful evidence in effective assessment. Too often it is assumed that for assessment to be effective it must use only numbers, not narratives. In reality, there is a story behind most every statistic, and some stories simply cannot be told well with statistics alone. Some schools may have the mistaken impression that if assessment results are expressed quantitatively (e.g., our students ranked this at 4.5 out of 5), then somehow those results are more reliable or more objective.

*Even the act of gathering quantitative data is inherently subjective: what data are gathered, how they are gathered, and from whom they are gathered are all subjective decisions. This caveat is not meant to diminish the value of quantitative data but to emphasize the equally important role that qualitative data can play. Effective assessment plans have both.*

Others get so excited about numeric data that it overshadows the point of the data itself, which is to serve the goals of analysis and improvement.

To be sure, quantitative data can be very useful and should be part of any school's assessment plan, but it is a bit of a fallacy to assume that because assessment results are expressed numerically, they are not also subjective. All results are subject to interpretation (see 5.c below). In fact, without interpretation, assessment results are fairly useless. Even the act of gathering quantitative data is inherently subjective: *what* data are gathered, *how* they are gathered, and *from whom* they are gathered are all subjective decisions. This caveat is not meant to diminish the value of quantitative data but to emphasize the equally important role that qualitative data can play. Effective assessment plans have both. Evaluation committees who visit member schools want to see, for example, some statistical data about how well students rate their educational experiences, but they also want to interview students and hear firsthand the nuances that numbers can't communicate. For further discussion about using numbers in assessment, see Appendix 2: Excursus on Quantitative Data.

Effective assessment recognizes the interrelated nature of quantitative and qualitative data. For example, a school might conduct a qualitative review of masters' theses but express those qualitative results in quantitative terms. In this example, a school might conclude that when faculty used an agreed-upon rubric to conduct a qualitative review of 10 randomly chosen MA theses written over the past five years, it found that 40 percent demonstrated achievement of all four MA student learning outcomes at the highest (4) level, while the remaining 60 percent achieved all four outcomes at an acceptable (3) level.

### Direct and indirect measures

Direct and indirect measures are also both important and useful ways to gather evidence for effective assessment—and either can be expressed in quantitative or qualitative ways. The simplest distinction between these two is this: direct measures assess *student performance*, while indirect measures assess *student perceptions*. For example, a juried music recital is a direct measure of a student's performance, while a survey of students about the educational effectiveness of a music class is an indirect measure of students' perceptions. Most ATS

*Direct measures assess student performance, while indirect measures assess student perceptions.*

member schools tend to have a wealth of indirect measures of student perceptions, the most common being course evaluation results. However, Educational Standard, section ES.6.3.2, “requires *both direct* (performance based) *and indirect* (perception based) *measures* of student learning” (emphasis added). Assessment based only on indirect measures runs the risk of not being able to document *what* (or *that*) students actually learned, only what they *think* they learned. On the other hand, indirect measures help a school understand the level of student satisfaction that exists—and in this consumer-oriented higher education landscape a school not concerned about student satisfaction runs the risk of losing students. Indirect measures also value the opinions and expertise that students, as adult learners, bring to their own learning. They are often an appropriate judge of whether a particular learning goal was met, even if they ought not be taken as the only judge. For example, both the ATS Graduating Student Questionnaire (GSQ) and Alumni/ae Questionnaire (AQ) are examples of indirect measures, largely in the form of questions about student satisfaction. These indirect measures can provide very helpful information to a school but become stronger when paired with direct measures, such as artifacts and observations.

For more information on these and other ATS assessment instruments (e.g., the Institutional Peer Profile Report [IPPR] and the Strategic Information Report [SIR]), please visit the ATS website under [Accrediting > Self Study and Assessment](#). The use of these instruments by ATS member schools is strictly voluntary. Using these instruments does not guarantee a positive accreditation decision, just as not using them does not necessarily lead to a negative decision. As noted at the outset, since there is no single best way to do assessment, these ATS resources are simply tools that a school might (or might not) choose to use; and, as with any tools, they can be used well or poorly. For schools that do choose to use the ATS Q's (a collective reference to the Entering Student Questionnaire, the Graduating Student Questionnaire, and the Alumni/ae Questionnaire), a very helpful guide for how various items from those questionnaires relate to different Commission Standards can be found on the ATS website under [Resources > Student Data > Resources for Using the Questionnaires](#). Archived report results for the Qs from all ATS member schools who choose to use those instruments can also be found under [Resources > Student Data](#).

Schools sometimes struggle with finding appropriate direct measures. Unlike music schools or law schools, theological schools do not have a tradition of requiring juried recitals or bar exams. And unlike many professional schools, there are no nationally normed tests for theology or ministry like there are for, say, psychology or dentistry. Even within denominations that have qualifying examinations or other comprehensive ordination processes, the outcomes assessed there may differ significantly from the school's program goals or outcomes or may attend only to a few of them.

*Perhaps the most common type of direct measure is a culminating experience of some sort.*

However, there are many different ways that ATS member schools utilize appropriate direct measures. We know of one school, for example, that actually does have "senior recitals." This school has graduating students "perform" in a scheduled event where faculty and other professionals observe them as they speak about current events, perform role-plays based on case studies, plan a strategy to address a difficult issue in a parish, and even compete in a theological "quiz bowl." These "performances" are directly linked to each of the school's degree goals or outcomes. Other schools may schedule a weekend retreat near the end of the final year during which students interact with faculty and other professionals in a variety of experiences that demonstrate how well they have achieved the goals or outcomes for their program, including oral interviews, mock lessons or even counseling sessions, and written or oral responses to various case studies.

Perhaps the most common type of direct measure is a culminating experience of some sort, such as a capstone course or summative project (MA thesis, DMin project, PhD dissertation or exam, etc.). In some capstone courses, the course objectives are simply the degree program goals or learning outcomes, with various assignments demonstrating the extent to which students have achieved each. For example, for a degree program preparing students for further graduate study, the assignments in such a course might include a research paper (demonstrating desired learning outcomes related to writing skills, research capacity, and scholarly voice), an annotated bibliography (demonstrating familiarity and fluency within a scholarly discipline), and a reflection paper that discusses the student's journey through the degree program and goals for the future (demonstrating, even if somewhat indirectly, how a student's outlook has changed as a result of completing the degree).

Supervised ministry experiences can also provide very helpful direct measures, by asking supervisors to evaluate various student “performances,” such as sermons preached, lessons taught, calls made, meetings facilitated, projects completed, and so forth. Again, those results are aggregated anonymously for faculty review to see how well the program is performing, not simply how well each student is performing. Sometimes the most telling data are the simplest: would you hire this student? For some schools, one of the most helpful measures is how students do on the denomination’s ordination exam. Those data can often be compared internally with previous classes to show trends and compared externally with other denominational seminaries to show areas of relative strength and concern.

One helpful way to enable and emphasize the use of direct measures is through student portfolios. In fact, in just the last five years, the percentage of schools—at least in the United States—using portfolios to assess student learning has skyrocketed from 5 percent to 45 percent, based on a survey of chief academic officers conducted by the National Institute for Learning Outcomes Assessment (see their March 2014 report, [Institutional Assessment Practices Across Accreditation Regions](#)). The advantage of using student portfolios is that rarely can a single assignment or activity adequately demonstrate how well the students as a whole in a degree program meet that program’s goals or outcomes, particularly in professional degree programs.

Portfolios are collections of key assignments (or artifacts) scattered throughout the curriculum that faculty decide are representative of how well each degree program goal or student learning outcome has been achieved by students in that program. Many portfolios are now collected electronically, especially in schools with learning management systems that allow students to submit selected assignments online at various points throughout their programs. Faculty later then assess (or discuss the original assessment of) those selected assignments collectively (usually without any student names) and aggregate the results, because the *purpose is to see how well all students are achieving* those program goals or outcomes, *not just how well each individual student is achieving* them.

As with any other assessment strategy, an important caveat to remember with portfolios is that they need to be closely connected to degree program goals or student learning outcomes (not simply the aggregate

of all course objectives) and to an assessment plan that is simple, sustainable, and sufficient. It is not necessary, or usually helpful, to collect every assignment that every student produces or to redo the work of the original instructor in terms of grading and feedback. A portfolio system can be quite helpful, though, when specific assignments are linked to specific learning goals (e.g., exegesis papers that demonstrate students' abilities to perform exegesis) or when assignments can demonstrate learning goals beyond their disciplinary box (e.g., sermons that were originally assigned to evaluate students' skills in homiletics might also be useful as a way of seeing whether students use the texts of their traditions in effective ways).

### Benchmarks or performance indicators

*Each degree program or learning outcome needs some kind of "performance indicator" or "benchmark" for the faculty to interpret assessment results meaningfully.*

In using both quantitative and qualitative data through both direct and indirect measures, it is important for effective assessment to have some sense not only of what information is needed, but also of what "criteria for success" will be used to evaluate or interpret that information. While this discussion could be included under 5.c (Interpretation of information), we raise it here since it relates so closely to the information collected. Each degree program goal or learning outcome needs some kind of *performance indicator* or *benchmark* for the faculty to interpret assessment results meaningfully. Keep in mind that each goal or outcome may have more than one performance indicator or benchmark, especially for goals or outcomes that address complex issues. To be sure, the ATS Commission Standards do not use these terms. However, the *Educational Standard* does require faculty to interpret assessment results (see section ES.6.4.1). Consequently, faculty will need some way of interpreting the results in terms of whether that goal or outcome has been achieved. Many schools choose to use terms like *performance indicators* or *benchmarks* or *criteria for success* to help faculty interpret those results. For example, it is not enough to simply say that the students in this program averaged 3.5 on a 5.0 scale or rubric for a given goal or outcome. What does that 3.5 mean? Is that good or bad? Are we pleased or displeased? Are results improving or declining? Is an average the best way to assess this goal/outcome? Or should some other factors be considered (e.g., the entire group will average 3.5 or at least 80 percent of the group will score at least a 3.5)?



*To ensure that any benchmark tied to a graded assignment is closely linked to learning goals and is applied fairly and uniformly over time . . . the faculty should develop a clear grading rubric.*

Here is an example of how a performance indicator or benchmark helps assess a given goal or outcome. A seminary has as one of its MDiv program goals or outcomes: “Graduates proclaim Scripture with appropriate attention to the ancient context of the text and to the current context of their audience.” The seminary may use a key assignment in an exegetical course (course-embedded assessment) to measure how students achieve that goal or outcome. However, having a goal or outcome and a measure are not enough. There must also be some benchmark against which to measure whether that goal or outcome has truly been achieved through that assignment. Such a benchmark is often called a performance indicator. For example, a performance indicator or benchmark for the previously cited goal or outcome might be this: At least 85 percent of the students completing the exegetical project for Course X will receive a score of at least 3 out of 4 on a rubric for that assignment (see rubrics below). That indicator or benchmark uses both quantitative data (at least 85%, which specifies a quantity of students) and qualitative data (score of at least 3 out of 4, which attests to the quality of student projects) for this direct measure of student performance.

As noted earlier with goals or outcomes, performance levels should be set at reasonable rates. For example, to set the levels at 100 percent for every goal or outcome is typically too idealistic because the failure of a few students to achieve a given goal or outcome does not mean that the entire program is ineffective. On the other hand, setting the level at, say, only 50 percent may be too easy a target to hit. There is no magic number, only a reasonable figure developed through appropriate conversations with key constituencies for the purpose of providing useful information that might then inform efforts for improvement or change.

## Rubrics

It is important at this point to highlight the value of grading or scoring rubrics. Rubrics are simply guidelines for rating student performance and one of the fastest-growing components in effective assessment. The same March 2014 NILOA report noted earlier indicates that rubrics have increased in use over the last five years from an average of less than 25 percent to nearly 70 percent among institutions of higher education—at least in the United States. To ensure that any benchmark or performance indicator tied to a graded assignment is

closely linked to learning goals and is applied fairly and uniformly over time (a key component of effective assessment), those involved in grading that exegetical project should develop a clear grading rubric. That rubric clarifies what is expected of students to achieve a B, or an A, or what would constitute a C or less. Grading rubrics for course-embedded assignments should also be tested by other qualified faculty (or even external parties appropriately qualified) to minimize the often arbitrary nature of course grades and to ensure connection to the overall degree learning goals. And some rubrics do not relate to course grades at all but are used instead to help faculty evaluate a goal or outcome that is not linked to a specific course assignment. For further discussion of rubrics, see Appendix 3: Excursus on Rubrics.

### Demography and delivery

Before we leave this discussion of data collection, it is worth mentioning two other ways that data may be collected and categorized, namely by demography and by delivery. Some schools, for example, may find it very helpful to distinguish assessment data by gender, age, or race/ethnicity—issues of demography. Such data can prove very useful in devising strategies to improve student learning, since some strategies may be more effective with certain demographic groups.

Additionally, some schools may find it very helpful to distinguish assessment data in terms of delivery systems. Schools, for example, offering MDiv programs through onsite, offsite, and online modalities may want to compare what students are learning through each of those delivery systems. As with demography, the effectiveness of certain strategies to improve student learning may depend on the delivery method utilized. Such distinctions can be particularly valuable in assessing factors like personal and spiritual formation, one of the four content areas required for MDiv and professional MA degree programs. Distinguishing assessment data in these ways does not mean schools need to use totally different assessment approaches. It just may mean some schools will use the same collection instruments but disaggregate those data by demography or delivery. Having that kind of information available to faculty, and other key players, can assist in the interpretation process—the subject of the next part of this guide.

### 5.c Interpretation of that information by key players—especially faculty

As we noted earlier under 4.a (Is it simple?), if key players, particularly the faculty as a whole, are not seriously engaged in assessment with meaningful conversations about the results, then the assessment plan is pointless and the results are useless. That may seem like an overstatement, but the ATS Commission Standards make it abundantly clear that “the faculty” (as a whole) play the primary role in using assessment results to improve student learning. Educational Standard, section ES.6.4.1, puts it rather pointedly:

Evaluation of student learning is the **responsibility of the faculty**. Effective assessment of student learning *requires that the faculty review and analyze the results of the assessment activities, discern appropriate changes to curriculum and educational practices, and document the results of both its interpretation of assessment data and decisions about educational changes* (emphasis added).

*The ATS Commission Standards place such a high priority on faculty involvement in assessment because they place such a high priority on faculty in the student learning process.*

Too many assessment reports reviewed by the ATS Board of Commissioners “fail the faculty test” because there is no evidence that faculty are seriously engaged in the assessment process, at least in interpreting that information. The ATS Commission Standards place such a high priority on faculty involvement in assessment because they place such a high priority on faculty in the student learning process. If one did a Wordle™ display of the Standards, three words would dominate the graphic: theological, faculty, and students. The Standards recognize the invaluable and irreplaceable role that faculty play in communicating theological truths to students. That role is no less vital in assessment, because assessment helps a school understand how well it is achieving its mission, particularly its educational mission of enhancing student learning. While some aspects of assessment can be delegated (primarily data collection), the interpretation of assessment information rests best with the faculty who know best the aims of theological education.

Another reason why the ATS Commission Standards prioritize faculty involvement in interpreting assessment results is that faculty have the best understanding of what students should be learning. It is, after all, their curriculum. As Standard 7 on authority and governance

*Assessment data can lead to poor decisions if not interpreted correctly and contextually.*

(section 7.3.3.1) states: “Within the overall structure of governance of the school, authority over certain functions shall be delegated to the faculty . . . [including] oversight of academic and curricular programs and decisions.” Assessment data can lead to poor decisions if not interpreted correctly and contextually. For example, one person might see a low score on a particular program goal and conclude that something is wrong and must be fixed immediately. However, faculty involved in that degree program might well understand that the particular score in question was the result of a nonrepresentative sample (“bad classes” do happen) or was the result of a one-time experiment (since abandoned) or was due to a fault of the assessment tool (some tools may be valid, but the results are not reliable; others may have reliable results, but the tool itself does not validly measure what it is intended to measure). Anomalies do not portend a trend, and poor data can lead to poor decisions. Healthy conversations among faculty and with administrators can clarify complex data that at first seem so simple. False negatives (and false positives) can occur in assessment as much as in any kind of data-based research.

One other reason to give faculty such significant responsibility in interpreting assessment information is that the faculty are the ones who will need to make any resulting changes to improve student learning. While administrators or trustees may see what needs to be done, shared governance in a community of faith and learning calls for conversations among faculty so that any changes can be implemented with an appropriate sense of faculty ownership.

Before moving to another key player in the interpretation of assessment information, it is worth suggesting ways in which faculty can be involved without being overwhelmed. As noted above, some ATS member schools have found it helpful to delegate some aspects of assessment to other parties, such as to an administrator or staff person or even a single faculty member. That person or office typically coordinates the overall assessment plan and process, especially the collection of assessment information, which is then provided to faculty in a synthesized, manageable form. Faculty as a whole then spend their time interpreting the resulting assessment information, not in data collection, though using faculty time to evaluate key student artifacts is also quite valuable. Many ATS member schools have an “assessment day” or “assessment retreat” either at the end or the beginning of the

*“The governing board of the school is responsible for ensuring that the school has a program of assessment of student learning and that the results of these activities inform and shape educational and institutional decisions.” (ES.6.4.3)*

academic year. That is often a day or two of focused faculty time spent on evaluating or interpreting assessment data and sometimes also spent on reviewing artifacts of student learning.

While faculty play a vital role in interpreting assessment data (and using those interpretations to effect meaningful change, as we discuss in 5.d below), there is still one other key player in this process. As noted earlier, the current ATS Commission Standards, approved by the membership in 2012, make it clear that “the assessment buck stops” in two places: with the faculty and with the board. To cite again the ATS Commission’s Educational Standard, section ES.6.4.3: “The governing board of the school is responsible for ensuring that the school has a program of assessment of student learning and that the results of these activities inform and shape educational and institutional decisions.” A very effective way to do that is *through* the faculty, but perhaps an even more effective way might be *with* the faculty.

Some member schools, for example, involve trustees in faculty conversations about how the assessment information is interpreted. If trustees represent a school’s various constituencies, especially churches, having their input in this conversation can be a helpful “external review.” This suggestion is not meant to blur the lines between governing and teaching, but it is meant to acknowledge the vital voice that many trustees bring to the educational table. And schools who do have such engagement between faculty and trustees over assessment of student learning find renewed meaning in the phrase “shared governance” and renewed appreciation for what faculty do, day in and day out. Faculty also have a deeper appreciation for the important role that trustees play—beyond the typical financial and business aspects of the board. It allows both key players to focus on what matters most—mission.

### **5.d Implementation of those interpretations for meaningful improvement**

When all is said and done about assessment, not answering well the “So what?” question makes the entire assessment process devolve into much ado about nothing. If assessment does not lead to improvement, the process seems pointless. Otherwise, the common faculty critique that assessment is just about satisfying some accreditation regulation is a valid one. Remember that accreditors, however, are not looking

*Remember that accreditors are not looking primarily for paperwork; they are looking for improvement.*

primarily for paperwork; they are looking for improvement. One of the most common questions that peer accreditors ask when they are on your campus evaluating your assessment efforts is this: What difference has this made for your students and school? They will often ask teachers and trustees to cite a few examples of how assessment led to improvement in student learning. Regardless of how elaborate the process and extensive the paperwork, if no one can readily list key changes that came from assessment, then that assessment is not effective.

To be sure, often the best interpretation of assessment information is to “stay the course,” meaning that a school is doing some things very well and should not change those things. In such cases, the right interpretation is a correct confirmation that improvements have already occurred and should simply be maintained. The ATS Board of Commissioners, however, has yet to review a school that did not need to improve in *some* area. There is no perfect seminary. And appropriate interpretation helps a school implement those changes that are truly needed and not waste its time changing just for the sake of change—or not changing at all while their educational equivalent of Rome burns around them.

In short, assessment is still about improvement. The first three components in the assessment cycle mean very little if schools don’t “close the loop” by using the information to implement meaningful changes. Early on in the assessment movement, T. J. Marchese made that connection clear with this memorable metaphor: “Assessment per se guarantees nothing by way of improvement, no more than a thermometer cures a fever” (from “Third Down, Ten Years to Go,” *AAHE Bulletin*, 1987). Assessment has a thermometer function, but it is most effective as a thermostat. Here are some examples of meaningful improvements that have been implemented as a result of appropriately interpreted assessment information:

- A school improved significantly (from 3.2 to 3.8 on a 4.0 scoring rubric) its students’ ability to integrate theology and ministry when it revised a core course to target that goal and had two faculty from each area team teach it—a strategy suggested by several students in the assessment process. Faculty expressed strong affirmation of how students are now better integrating what they learn. To quote a typical response from one professor, “I would not have

thought that this one change could make such a big difference, but our students' ability to integrate theology and ministry seems so much stronger now."

- A school improved by 15 percentage points its students' preaching skills (from 20% rated excellent by the faculty on a scoring rubric to 35%) when it used assessment data from its various preaching courses to redesign its MDiv to include preaching assignments throughout the curriculum, rather than in isolated courses. Assessment information from church leaders has confirmed the improvement noted among recent graduates. One pastor went so far as to state, "These students' improvement in preaching is nothing short of profound. Thank you."
- A school improved significantly the quality of its international students' final project (from an average grade of C+ to B+) when it used its assessment data to replace its faculty-taught writing and research course with a system of ongoing peer tutoring by student volunteers. This improvement is further demonstrated in the 25 percent increase in the number of advanced degrees being pursued by the school's international students.
- A school improved significantly its students' perception of their spiritual growth (from 3.1 to 4.3 on a 5.0 scale) when it used assessment information to supplement its faculty-led formation program with external mentors who meet weekly with students. Several students have commented that they have found "mentors for life" through this new approach.
- A school improved by 20 percent the placement rate for its MDiv degree program after instituting a capstone experience that involved denominational leaders and pastors in evaluating students' readiness for ministry—many of whom were then hired by those same leaders and pastors, who, in turn, have become more active recruiters for the school.
- A school improved significantly its completion/graduation rate for its MDiv degree program (from 65% to 90%) by redesigning its program content to make it more accessible and achievable, based on assessment data from students and in response to needs expressed by constituent churches. These improvements have



also helped the school address its long-standing financial concerns through increased student revenue.

We recall the line from *The Dead Poets Society*: “Poetry . . . reminds us that the powerful play goes on, and you may contribute a verse.” Assessment is not numbers on a Likert scale. Assessment of theological education is no less a challenge and no less a reward than poetry—more art than science. “We’re not laying pipe.” We’re talking about . . . people and purpose and passion—about mission and ministry. *That* is why assessment is important. What will your verse be?

## Section 6 A Checklist for Effective Assessment

This last section attempts to summarize and synthesize best practices of effective assessment in the form of a checklist, based on the “four foundations of effective assessment” above. This checklist can never be a substitute for the thoughtful and purposeful work that effective assessment requires. It is not meant to be a red-button “easy answer” to assessment. This checklist is simply a way of remembering key components of effective assessment—components that can form an effective assessment plan and process. Schools should use this as a starting point for further conversation, reflection, and action. This checklist also serves as a guide for peer reviewers and Commissioners in evaluating effective assessment.

Checklist for Effective Assessment				
<b>1. Effective assessment plan and process</b>	<b>Standard*</b>	<b>YES</b>	<b>NO</b>	<b>IN PROCESS</b>
a. Is appropriate mandate/oversight provided by the governing board?	ES.6.4.3			
b. Is there a designated person or office to coordinate assessment?	ES.6.2.3			
c. Does plan/process reflect school's mission, ethos, resources, and size?	ES.6.2.1,3			
d. Do faculty play a central role in the assessment process?	ES.6.4.1			
e. Is the plan/process simple, sustainable, sufficient/adequate?	ES.6.1			
f. Is the plan/process itself assessed and changed as needed?	ES.6.2.4			
Committee Comments:				
<b>2. Identification of desired goals/outcomes—for each degree</b>	<b>Standard*</b>	<b>YES</b>	<b>NO</b>	<b>IN PROCESS</b>
a. Are goals/outcomes clearly stated for each degree program?	ES.6.1, 6.4.2			
b. Do goals/outcomes reflect appropriate Degree Program Standards?	A-J.1.2**			
c. Are goals/outcomes expressed in demonstrable terms?	ES.6.3.1,2			
d. Are goals/outcomes manageable in number?	ES.6.1			
e. Are goals/outcomes decided by faculty in those programs?	GIS 5.3.1***			
Committee Comments:				

3. Information on the achievement of those goals/outcomes	Standard*	YES	NO	IN PROCESS
a. Does information come from both direct and indirect measures?	ES.6.3.2			
b. Does information include both quantitative and qualitative data?	ES.6.3.1			
c. Does information use some form of benchmarks?	ES.6.1(2)			
d. Does information include data about completions/ placement?	ES.6			
e. Does information not rely mostly on course grades; use rubrics?	ES.6.3			
f. Is information about individual students kept confidential?	ES.6.3.3			
Committee Comments:				
4. Interpretation of that information by key players, especially faculty	Standard*	YES	NO	IN PROCESS
a. Are the key players, especially faculty, engaged in assessment?	ES.6.4.1			
b. Do faculty have substantive conversations about assessment data?	ES.6.1(3)			
c. Do faculty review regularly the implications of this information?	ES.6.1(1)			
d. Do faculty interpret raw data in appropriate ways in light of mission?	ES.6.2.1			
Committee Comments:				
5. Implementation of those interpretations for meaningful improvement	Standard*	YES	NO	IN PROCESS
a. Do faculty interpretations lead to improvements in student learning?	ES.6.4.1			
b. Does school provide clear examples of such improvements?	ES.6			
c. Does board ensure results shape educational/institutional decisions?	ES.6.4.3			
d. Are assessment plan/process and results clearly documented?	ES.6			
e. Does school summarize key assessment results publicly and regularly?	ES.4.4			
Committee Comments:				
* Refers to the ATS Commission's <i>Educational Standard</i> (ES), unless otherwise noted below.				
** Refers to section 1.2 of each Degree Program Standard (A-J).				
*** Refers to the ATS Commission's General Institutional Standard 5 on faculty.				

## Appendix 1

### Sample Summaries of Assessment Results

NOTE: The samples in this appendix are meant to be descriptive, not prescriptive—to provide illustrations, not mandate models. Either of these examples might well form part of an assessment report (e.g., an appendix), but not the entire report. They need to be complemented with a narrative that provides context and nuance, including discussion about what worked well and what did not, what improvements have been made, and what still needs to be done.

#### Sample 1: Assessment Results Summary for an MDiv Chart Format Using Nichols Model (see 4.a)

Student Learning Outcomes	Direct and Indirect Measures	Criteria for Success (Benchmarks)	Assessment Results*	Changes Made (as appropriate)
1. Graduates will present Scripture with depth and in ways consistent with the school's statement of faith.  (Standard A.2.2)	<i>Direct Measures</i> Entering Bible exam Exit Bible exam Exegesis Paper in BL702 Juried Review Sr. Sermon <i>Indirect Measures</i> ATS GSQ Table 13 ATS GSQ Table 23 Annual Alumni Survey	65% score at least 65% 85% score at least 85% 85% score 3–4 on rubric 100% rated “above avg.” Average rating of 4 of 5 Average rating of 4 of 5 Average rating of 3 of 4	70% scored 65%+ 88% scored 85%+ 80% scored 3 or 4 90% rated above avg. Averaged 3.5 of 5 Averaged 4.5 of 5 Averaged 3.2 of 4	No changes; met No changes; met Added earlier paper 1 student “anomaly” (so no change) Added earlier paper No changes; met No changes; met
2. Graduates will manifest global awareness and cultural sensitivity, demonstrated through selected assignments.  (Standard A.2.3)	<i>Direct Measures</i> Global engagement project  Cultural awareness scale  <i>Indirect Measures</i> Survey on global and cultural expressions	80% “meet expectation” on project rubric 75% score “above average” on scale  Graduates average 3.5 out of 4 on survey	50% met expectations 60% scored above average  Grads averaged 2.5 out of 4	None of the benchmarks for this outcome was achieved; dean appointed faculty task force to make recommendations for curricular change
3. Graduates will demonstrate mature spirituality as determined by faculty and field mentors, as well as through self-perception.  (Standard A.2.4)	<i>Direct Measures</i> Field mentor evaluation  Faculty advisor evaluation  <i>Indirect Measures</i> Exit survey/interview  Annual Alumni Survey  ATS GSQ Table 19	Mentor rating of “meets or exceeds expectations” Faculty rating of “meets or exceeds expectations”  Average rating of 4 of 5 Average rating of 4 of 5 Average rating of 4 of 5	No mentor ratings were collected Faculty ratings met that benchmark  Averaged 3.9 of 5 Averaged 4.1 of 5 Averaged 3.9 of 5	Added mentor training  No changes; met  Changed benchmark (grads too “humble”) No changes; met (alum noted “humility”) Appointed faculty task force to review
4. Graduates will ...	NOTE: This outcome has been deleted for the sake of brevity.			
* The assessment results listed here are simplified samples. An actual assessment report should specify the time period covered (e.g., 2014 graduates), as well as give some indication of the total population (e.g., 15 total graduates from 2014) and the number/percentage of those providing assessment data (e.g., data from 12 of 15 were collected), and any sampling methods used (e.g., random sample). For more discussion, see the third point in Appendix 2.				

## Sample 2: Assessment Results Summary for an MDiv Narrative Outline Format

### Outcome 1: MDiv graduates will demonstrate appropriate skills in interpreting Scripture.\*

#### Outcome Measures Used

1. Direct Measure: Exegesis paper from MDiv portfolio.
2. Direct Measure: Capstone research project, scored using faculty-wide rubric developed last year.
3. Indirect Measure: ATS's Graduating Student Questionnaire (GSQ) on Table 13, which has two relevant questions regarding how well students feel prepared to use and interpret Scripture, and preach well.
4. Indirect Measure: Course evaluation forms for all biblical courses averaged over time, using a rolling average of the last five years (the average time to complete an MDiv).
5. Indirect Measure: Exit interviews from MDiv graduates.

#### Criteria for Success

[NOTE: "Criteria for Success" = "Performance Indicator" or "Benchmark"]

1. Average score of "acceptable" on rubric for exegetical paper in MDiv portfolio, with the added goal of having at least one-fifth (20%) achieve an "exceptional" rating.
2. Average score of 3 ("acceptable") out of 4 on the capstone project rubric for all MDiv grads, which includes the ability to interpret and apply two different texts in case studies.
3. 80 percent of graduating students taking the GSQ will report a score of 4.0 or above (out of 5)
4. 75 percent of all students will rate their Bible courses at 4.0 or above (out of 5), using the internally developed course evaluation form in use over the last 10 years.
5. A random sample of exit interviews will generally be positive, with affirmations and helpful suggestions far outweighing concerns or complaints.

#### Assessment Results

1. MDiv grads averaged above "acceptable" on rubric for this artifact in the MDiv portfolio, with 30 percent scored at the "excellent" level.
2. MDiv grads averaged 3.5 on capstone rubric; 1 of 20 grads rated 1.0 (unacceptable), with that one student viewed as an "outlier."
3. 95 percent of MDiv grads have self-rated score of 4 or above on those questions, with 4.35 as the average, based on results from last year (GSQ is used every other year).
4. 73 percent of all students rated their bible courses at 4.0 or above (with 22% rating below 3.0), with an average response rate of only 58 percent over the last five years.
5. Appendix A provides a summary of student comments from last year's MDiv exit interviews, with comments providing strong affirmation of this outcome.

### Assessment Changes

1. Faculty discussed this at length and concluded that this is a strong area for our MDiv grads; they also recommended that the rubric (introduced three years ago) be revised with clearer criteria.
2. Faculty also felt the data indicate that students do well in achieving the MDiv program's first outcome on interpreting Scripture; the failure of one student to score an acceptable rating on the capstone was indicative of larger issues with this student, who has since been counseled out of the MDiv program.
3. The 95 percent rating reinforces the faculty's sense of this outcome as a strength of the program, though two of the current five faculty in Bible plan to retire in two years, creating some concern.
4. The failure to achieve this benchmark was a subject of much faculty discussion, with various reasons given. While the benchmark was almost met, the low rating by more than one-fifth of the students (22%) merits further discussion. A faculty task force was appointed by the dean last fall to bring back recommendations.
5. Student testimonials were very encouraging and quite positive, so no changes are anticipated.

\*NOTE: Only one MDiv outcome listed here for sake of space, but all MDiv degree programs must address the four program content areas described in Degree Program Standard A, section A.2.

## Appendix 2

### Excursus on Quantitative Data

In Chapter 5 (5.b), we discuss the use of quantitative and qualitative data. It may be helpful here to say a few more words about numbers.

*First*, schools often wonder how big a number they need to have adequate assessment results. Some might argue, for example, that since they graduate only a handful of students each year, then that number is too small to yield meaningful assessment results. In reality, small numbers mean a school can evaluate the entire population and not have to rely on a sample. Assessing what all five of your graduates have learned is just as meaningful, if not more so, than assessing what 25 of 30 graduates have learned. However, small numbers can also present a challenge for some schools. For example, a school assessing a group of only a few graduates might rightly be concerned that the data from one person could significantly skew the results. In such cases, several options are possible: (1) report only frequencies, not averages (e.g., four students ranked this outcome at 4.5 out of 5, while one ranked it at 3.0); (2) aggregate several groups of graduates so the results from just one person won't be so dramatic; and/or (3) report the results from the small group, but note in the report that one student represented results significantly different from all the others (a “statistical outlier”).

*Second* (and conversely), that does not mean that schools with large numbers of students have to assess every student every year in every area. Smaller samples of larger numbers can be a very efficient and equally effective way to conduct assessment—as long as you make sure that the sampling is legitimate. For example, a school might choose to have faculty review 10 out of 30 student assignments as an indicator of how that assignment achieved a specific student learning outcome for a specific degree program. Evaluating just 10 out of 30 (or 5 out of 12; there is no required ratio) is perfectly legitimate as long as those samples are not inappropriately chosen, such as choosing assignments only from known “A” students. It is also a good practice to use a random or representative sample where the identities of the students are not disclosed to the reviewers. The Standards expect schools to “guard the confidentiality of student work used in the assessment of student learning and, as necessary, provide for the anonymity of student identity in certain artifacts of their work” (ES.6.3.)

*Third*, if a school samples a larger audience or reports responses from less than the entire population of those being assessed, it is always best to give some indication of the total population, the sample size, the response rate, and/or the sampling method. For example, 75 percent of our 40 graduates last year indicated . . . Or, our faculty reviewed 10 out of 30 exegetical artifacts, chosen randomly. Or, among the 50 alumni/ae surveyed, only 15 (30%) responded, indicating that the results may not be representative of our graduates as a whole. And if a school is reporting on the whole population, it is still helpful to indicate the total number. For example, to report that 25 percent of our graduates “failed to meet expectations” (using the rubric for outcome X) may mean that 25 students “failed”—if



the total population is 100; but it may mean that only one student “failed”—if the total population is only four. Those results suggest two rather different responses.

*Fourth*, when reporting numeric assessment data, it is common to use simple descriptive statistics (e.g., mean, median, mode, frequency, etc.), not more sophisticated inferential statistics (Chi-square or t-tests, linear regression analysis, etc.). Anything more than simple descriptive statistics is usually not necessary. While there are some formulae proposed in statistical studies (e.g., for random sampling or size of sample), such sophisticated methods tend to be beyond the scope of most schools of theology. That is one reason why schools should usually report numeric data in not unduly specific ways (e.g., 77%, not 76.85% or 4.5 out of 5, not 4.475 out of 5). Rarely does the size of the populations represented among member schools merit anything more specific. Using overly specific numbers seems to imply that there is a significant difference between, say, a score of 4.497 and 4.505, when rarely there is.

## Appendix 3

### Excursus on Rubrics

In Chapter 5 (5.b), we discuss the value of rubrics. Some rubrics are used for specific course-imbedded assessments (e.g., exegesis paper) in order to bring greater clarity and consistency to their assessment. Other rubrics are used for assessing artifacts that are not linked to a particular course or assignment but are more broadly focused on overall student achievement in a degree program (e.g., capstone project or thesis). This appendix provides some further discussion and gives a few examples of both.

One of the values of using rubrics for specific course-imbedded assessments is to provide consistency and to combat the criticism of the arbitrariness of course grades. For example, if a professor grades a group of five papers and gives one A, three Bs, and one C using a well-developed grading rubric for that paper, then any other qualified personnel using the same grading rubric should assign virtually the same grades. A well-developed grading rubric helps ensure that any differences in grading are the result of student learning, not faculty arbitrariness. An additional advantage of grading rubrics is that students clearly know what is expected of them. If they do not meet appropriate criteria, then they know why and what they must do to improve—an important goal in assessment. Such rubrics can also save faculty time because they are not writing the same comment over and over again on dozens of different assignments.

It is worth noting at this point that rubrics, like assessment, should be simple. Effective rubrics typically have at most only three or four levels, such as excellent, acceptable, needs improvement, and/or unacceptable. They also should provide clear criteria for how one determines why something receives the score that it does. Elaborate 10-point scales that we see with some rubrics tend to make the process overly cumbersome and difficult to implement. Here are two examples of rubrics that are simple and clearly define the criteria by which the outcome is evaluated.

Sample Rubric for Thesis			
Outcome	Excellent (3.0)	Acceptable (2.0)	Needs Improvement (1.0)
Demonstrates appropriate writing skills	Has clear thesis, material is organized very well, no grammatical errors, presents more than one side of an argument	Has fairly clear thesis, material is organized, minimal grammatical errors, sometimes presents only one side	Thesis not clear or lacking, material poorly organized, many grammatical errors, presents only one side of an argument
Demonstrates graduate-level research skills	Uses at least 75 resources, resources represent best scholarship, resources are used professionally, resources are formatted consistently	Uses at least 50 resources, resources represent good scholarship, resources are used unevenly, resources sometimes not formatted consistently	Uses fewer than 20 resources, resources not from reputable sources, resources rarely cited, resources not formatted consistently
NOTE: A thesis will clearly have many other outcomes; these two are simply illustrative.			

Sample Rubric for Sermon			
Outcome	Exceeds Expectations	Meets Expectations	Does Not Meet Expectations
Grasps the attention of the audience from the outset	Memorable opening line that captures attention and sets up the topic/text	Clear opening line that fits well with sermon	Lacks memorable or clear opening; just starts with text or topic
States clearly the “big idea” of the sermon	Key point(s) of sermon are clear and easy to remember	Key point(s) of sermon is(are) fairly clear	Not clear what sermon is about
Demonstrates sound exegesis of appropriate text(s)	Uses Scripture well with clear and compelling interpretations	Interprets Scripture fairly, but with little explanation	Paid only “lip service” to Scripture with no explanation or interpretation
Uses language and illustrations appropriate to audience	Spoke clearly and correctly with powerful illustrations to make his or her point(s)	Spoke fairly clearly with only a few grammatical mistakes; used only a few illustrations	Was hard to understand, with numerous grammatical mistakes; very few, if any, illustrations.
NOTE: A sermon may have other outcomes; these four are simply illustrative.			

When used for degree program assessment, rubrics for a single course-embedded assignment should focus attention on the desired goals or outcomes of the program as much as possible. For example, a C paper may be poorly written but still demonstrate a student’s proficiency with exegesis; a B paper may be beautifully written but be lacking as far as the goal or outcome is concerned. This is another reason why course assignments and course grades, alone, are sometimes limited in their ability to serve the degree program assessment process.

For further information about developing rubrics, one could conduct a simple search on the Internet under “college-level grading rubrics” (“grading rubrics” alone will yield mostly results for elementary or secondary schools). Or more specifically, one could search under “rubrics for theology,” which yields some helpful examples, including some from ATS member schools. One could also visit the website for the National Institute for Learning Outcomes Assessment, which has an especially helpful guide for developing clear and concise rubrics.

We also call your attention to an assessment project being developed by the National Catholic Educational Association, with support from ATS. That project, still in process in 2014, focuses especially on using rubrics to assess the MDiv degree program in the context of the four pillars described in the Program for Priestly Formation for Roman Catholic seminaries in the United States.