A Reflective Guide to Effective Evaluation for Theological Schools

Approved by the ATS Board of Commissioners on June 7, 2021

ats The Association of Theological Schools
The Commission on Accrediting

Table of Contents

Introduction

Evaluation. It is a simple idea: to reflect thoughtfully on what we have done in the past and are doing in the present so that we can learn from these experiences and explore new possibilities in the future. Good evaluation can help us be better teachers, leaders, learners, pastors, and colleagues. It can help us achieve our goals and live out our callings; it helps us know what our missions are and how we might better achieve them. But evaluation is also a term and a practice that carries significant baggage: it is sometimes experienced as a waste of time, busywork, meddling; it triggers insecurities and defensiveness and imposter syndrome; it can be used to reinforce structures of bias or privilege; it can sap limited resources of money, energy, and time. It is a simple idea, but not at all a simple practice.

Evaluation has great potential for good and yet is often done badly. Perhaps that is why evaluation (including both institutional evaluation and the evaluation of student learning) is one of the most frequent areas of needed growth 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 institutional and educational evaluation can be among the most rewarding tasks that a school undertakes. This short guide is intended to highlight the high value that evaluation can have for your school, both in terms of evaluation of student learning and of institutional evaluation more broadly, and to share some practices that will help you engage this work more easily and more fully.

While we assume that all ATS institutions have some experience in evaluation, we also know that many individuals involved in evaluation are often new to their institution or new to their role, given the normal transitions among member schools. These transitions involve those who serve not only in offices of institutional effectiveness or as assessment coordinators but also as presidents and deans—and faculty and trustees. We mention the latter two especially because the ATS Commission Standards make it clear that "the buck stops" for evaluation in two places: with the faculty and with the board. Standard 9.5 notes that the governing body "ensures that the school's mission and educational and institutional outcomes are regularly evaluated and that the results are used to better achieve the school's mission and improve its various outcomes," and Standard 2.6 observes that faculty play a key role in the evaluation of educational outcomes.

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

This guide attempts to demystify evaluation in at least two specific ways. First, it *provides a concise and practical framework to help institutions imagine, implement, and improve their own*

evaluation plans and processes. Second, it *serves member schools by clearly linking the project of evaluation to the expectations of the ATS Commission Standards*.

As we begin, an important caveat is warranted. The ATS Board of Commissioners, made up of peers elected by the entire ATS Commission membership, believes there is *no single best way* to do evaluation. This has been true all along, and is especially so under the new 2020 Standards of Accreditation, which intentionally focus on *principles* of educational quality that schools can meet *in various ways*, rather than on a demarcated list of practices that each school must copy (as described in the *Preamble to Standards of Accreditation*). This means that schools can (and should) engage in evaluative practices that they find most helpful, in light of their own mission and context. What works well in one setting might well be burdensome, irrelevant, or counterproductive in another. Similarly, what once worked well for a school (e.g., when a degree program first started or had a large enrollment) might need to be changed as time goes on.

Yet this focus on principles, rather than best practices, does not mean that "anything goes." Schools are still required to implement evaluation plans and processes that are effective, with attention to both educational quality and institutional effectiveness, though the primary focus for any theological school is on students—"how well they are learning and how that learning helps them achieve appropriate personal and vocational goals" (Standard 2.8). The shift is something like going from a road map (which everyone must follow) to a directional compass (where we are all aiming for the same destination but might take different journeys to get there). While this can be freeing, some have observed that this focus on principles feels a bit like the guard rails are gone, and this might be especially true for the topic of evaluation, especially given that the literature on evaluation is extensive, even exhausting. This reflective guide seeks to help those who would like a few more landmarks on their journey.

We offer this brief guide not to break new ground but to highlight some good 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 evaluation,

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

but it is meant to be a helpful guide along the evaluation path, including the path that ATS member schools walk as they focus on evaluation of student learning within the context of graduate theological education. This document seeks to be illustrative and suggestive, not exhaustive. It is intended to help schools think about how to engage in evaluation more intentionally. In the end, every school's evaluative efforts will be evaluated in light of the Commission Standards, as interpreted first by the school in its context, next by a group of peer evaluators with their professional judgment, and finally by a Board of Commissioners who are elected by and act on behalf of the membership. We believe, however, that this guide will help schools better understand and implement the Standards regarding both

institutional evaluation and evaluation of student learning, and, more broadly, it can help your evaluative efforts be more effective, more efficient, and more engaging.

The title of this document, "**A Reflective Guide to Effective Evaluation for Theological Schools**," is intentional for these reasons:

- A *Reflective* Guide because it is more than just a fact-based set of instructions for doing evaluation. Rather, it contains practical and personal reflections regarding perceptions and practices of evaluation. As such, it is less formal than other documents produced for the Commission on Accrediting, using, for example, the editorial "we" throughout. This more personal approach is intentional in our goal of making evaluation more understandable to a wide range of constituents, including administrators, staff, faculty, and trustees.

- *A* Reflective *Guide* because it is intended as just that: "*a guide*" to effective evaluation, not as "the guide" or as "the only way" to do evaluation. The ATS Commission Standards of Accreditation, not this guide, constitute the final word on evaluation for member schools. As noted earlier, there are many good ways to engage evaluation, and not all of them can or will be included here. Still, schools that review this guide, based on those Standards, should be well positioned to meet the expectations on evaluation of the ATS Board of Commissioners.

- To *Effective Evaluation* because, while it does not (nor cannot) encapsulate every ideal of evaluation, it is based on good practices of evaluation recognized by the ATS Board of Commissioners to be effective; and it is founded on the ATS Commission Standards that frame and inform those best practices. This includes Standard 2 (Planning and Evaluation), but is not limited to Standard 2; evaluation is a theme that is woven throughout all of the Standards, is part of the guiding principles on which the Standards are built, and is central to the idea of accreditation as quality assurance and quality improvement itself (as will be discussed more fully later in this guide).

- *For **Theological Schools*** because this guide focuses on the work of evaluation from the perspective of the theological school as a whole. This includes three interrelated assumptions.

  1. As mentioned earlier, evaluation is not just the responsibility of a single role or department (e.g., office of institutional effectiveness, assessment coordinator, academic dean) but rather should be engaged by a wide range of stakeholders (including faculty, trustees, and administrators, among others). This guide will likely be helpful to those who have the word "evaluation" or "assessment" in their title, but it is intentionally not written only to them.

  2. While evaluation can and should happen at a number of different levels (from small to large), this guide will focus primarily on the sorts of evaluation that help a school understand how well it is fulfilling its mission, with particular attention on students—"how well they are learning and how that learning helps them achieve appropriate personal and vocational goals" (Standard 2.8). This guide

will likely be useful to those involved in any sort of evaluation process, but it is particularly focused on how evaluation serves a school's mission.

3. While evaluation is a helpful skill in almost any situation or process, this guide will focus primarily on evaluation for graduate theological schools. This allows us to lean into the depth of mission that is often grounded in faithfulness to a theological tradition and community, while also recognizing the organizational structures and institutional commitments that are part and parcel of graduate theological education. Some aspects of this guide might be helpful in other contexts—including that we believe evaluation and assessment are useful skills for our students to learn—but the focus is intentional on evaluation in graduate theological schools, and particularly those that are seeking to hold themselves accountable to the ATS Commission Standards of Accreditation.

This guide builds on an earlier Commission document, "A Reflective Guide for Effective Assessment of Student Learning" (originally published in 2014 as part of the Commission's *Self-Study Handbook*). The guide has been thoroughly revised and updated to align with the 2020 Standards of Accreditation and has been expanded to focus not only on evaluation of student learning (which we previously referred to as "assessment") but also institutional evaluation more broadly. We will discuss this shift more fully in the final section of Chapter One, both as it attends to the membership's desire for one unified set of standards incorporating institutional and educational evaluation together (see the annotated version of Standard 2.5) and that it recognizes that effective student learning cannot be separated from institutional effectiveness (and vice versa).

In addition to this guide, the Commission provides a variety of resources to help school with evaluation. Schools that are engaging in evaluation as part of the self-study process should review the Self-Study Handbook. Other resources are available on the ATS website; in particular, the section on Self-Study and Evaluation and the resources for Evaluation and Assessment Coordinators may be particularly relevant for schools developing or enhancing their evaluation plans, and the resources on Student Data, Institutional Data, and Data Visualization can be great supports for data collection and interpretation. The Commission makes a variety of educational and interpretive resources available to assist member schools, including workshops and online training resources related to evaluation, assessment, and data. And, of course, schools are reminded to be in conversation the Standards of Accreditation themselves, including the Preface, the Educational Principles (found in the Self-Study Handbook, pp. 5-7), and the Self-Study Ideas. Finally, each school is assigned a Commission staff liaison to support it in its work; your staff liaison would be pleased to be a conversation partner with you as you engage this significant work.

Chapter One: Philosophy of Evaluation ("Why?")

The Purpose and Importance of Evaluation

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" (COA *Bylaws*, Section 1.2). The Standards of Accreditation articulate the shared understandings and accrued wisdom of the ATS membership over many decades, while also attending to the diversity and variety of our schools today. As such, the standards reflect agreed-upon educational principles that help each member school better achieve its distinctive mission in light of its particular context. They assure the public of each school's educational quality—based on the professional judgment of peer and public members. They also foster flexibility and innovation. In all these ways, these standards help schools embody their missions, grow in light of their missions, and be transparent about their missions. It is the nearly 260 accredited members of the Commission that have developed and approved the Standards of Accreditation—and their significant focus on evaluation.

It is crucial to understand at the outset that *evaluation is not about compliance, but about a common commitment to continuous improvement* in our educational practices and our institutional missions. Evaluation is about improvement. To be sure, one oft-cited reason to care about evaluation is that both institutional evaluation (e.g., through a self-study process) and evaluation of student learning (e.g., through regular assessment events and reports) are required if an institution seeks to be accredited. And accreditation—either by the ATS Commission or by another body— is often either necessary or unavoidable, including that it serves as the gateway to U.S. federal financial aid (without which many of our schools, in both the U.S and

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

Canada, could not survive). However, to say that schools should engage in processes of evaluation because accreditation—or worse, the government—requires it is similar to faculty telling students to "learn this because it will be on the test." Evaluation 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?

Effective evaluation 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 evaluation 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. Evaluation is certainly not

> Evaluation 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.

about bowing to bureaucracy or being deluged by "administrivia" (as one dean once described it). Rather, evaluation 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.

Some Limits and Challenges to Evaluation

As important as evaluation is, it is only fair to name some of its limits and its challenges. The first limit we would list is that *evaluation 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 taking immediate actions to manage a financial crisis that they do not have time to focus on a comprehensive review of past practices. 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 evaluation is a key factor in addressing financial issues effectively, as it provides data that can help a school determine which expenses might be cut, which programs to expand, and which budget priorities to address. Put another way, evaluation itself can't put out the fire, but it can sometimes help a school identify which fire-suppression strategies to attempt.

Another limit to evaluation is often echoed in this familiar statement: *Not everything that is valued can be measured, and not everything that is measured should be valued*. Sometimes schools will begin an evaluation 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 backwards 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 evaluating theological education. Many of the most important things that ATS member schools care about cannot be measured by simple yardsticks, 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. Just because something is hard to evaluate 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. We are reminded of the man who lost his keys one night and was looking for them under a streetlight. 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." Evaluation needs to look where the keys might be, even if it's further away from the streetlight.

Related to this, we often hear from schools that one of the key challenges in institutional and educational evaluation is that *it can become an all-consuming, even overwhelming task*. To be sure, evaluation 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 *evaluating* 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." Evaluation is a terrific servant, but a terrible master. When portrayed as a master, evaluation processes can generate faculty resistance and be perceived as another unfunded mandate from accreditation bureaucrats. 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). Yet part of the joy and reward of teaching is having

> Evaluation is a terrific servant, but a terrible master.

meaningful faculty conversations about how well students are learning what faculty feel they need to learn—what is working and what needs attention (i.e., doing evaluation). 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. Evaluation of student learning 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).

Perhaps the most important limit and challenge to evaluation 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 evaluation plans that on paper seem to have every "jot and tittle" of what the books say evaluation should include, 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

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

accreditation requirement, with no chance of making any difference in those schools' purposes or people or programs. Such mechanistic approaches to evaluation 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 evaluation, 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." Evaluation is not numbers on a Likert scale. Evaluation 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 evaluation is important. What will your verse be?

Evaluation in the ATS Commission Standards of Accreditation

As mentioned in the Introduction to this guide, evaluation is a theme woven throughout the ATS Commission Standards approved by the membership in June 2020. This is apparent even from the opening **Preamble to the Standards**:

> Accreditation is about quality assurance for various publics and ongoing improvement for theological schools, especially regarding student learning and formation. It is a voluntary process through which schools mutually assure one another's educational quality with an eye toward ongoing improvement, based on standards. Through self-review, a school has regular opportunities to reflect intentionally on its distinctive strengths and its areas of desired growth in light of its unique mission and distinct context and in light of the standards. Self-review then supports the school's efforts in planning, evaluation, and imagination. Through peer review, an accredited school is endorsed by its peers as one of quality and integrity, which affirms the school's value to society, as well as its trustworthiness.

> Within the context of graduate theological education, accreditation is an ongoing way to live into the intersections of faith and learning. It involves giving close attention to the histories that ground us and the visions of the future that draw us forward. It is grounded in care for people, communities, and schools, now and in the future. It emphasizes stewardship and responsibility, while also holding space for grace and interdependence. It acknowledges the centrality of the unique mission of each individual school, while also recognizing that there is more that brings us together than separates us. Accreditation helps schools improve—not simply for their own sake, but primarily for the benefit of others, including the religious communities and other constituencies who serve and are served by their students. For all these reasons, accreditation is a deeply theological act with a focus on students, especially on student learning and formation.

Evaluation is also part of the foundation of the Standards themselves. For example, among the **seven membership priorities** articulated as part of the 2020 Redevelopment of the Standards, these three speak most clearly about the need for and value of evaluation (from the **Preamble to the Standards**):

> D. The standards ensure through evidence (qualitative and quantitative) that schools are effectively accomplishing their educational missions and continually seeking to improve in the achievement of those missions.
> E. The standards focus primarily on the quality of graduate theological education, attending to how well student learning and formation is achieved, however and wherever students are engaged.
> F. The standards focus on the health of both schools and the individual degree programs they offer.

A third place where this focus on evaluation can be found is in the **ten educational principles** in which the 2020 Standards are grounded. This is made explicit in principle #9, but can also be seen throughout all ten principles (found in the [Self-Study Handbook](#), pp. 5-7):

> 9. Theological education demonstrates careful institutional planning and evaluation. Graduate theological education builds from a clear sense of purpose, is undertaken through intentional processes of planning, is enacted through careful instructional and organizational design, and is evaluated in light of the mission and context of each school.

Finally, evaluation is woven throughout **the Standards themselves**. This is most apparent in Standards 2.5-8, under the heading of "evaluation":

> *Evaluation*
>
> **2.5** Evaluation is a process that engages appropriate constituencies to discern how well the various aspects of the school's mission are being achieved and how its educational and institutional outcomes could be maintained if met or improved if not met. Evaluation attends to all functions, personnel, and programs of the school. Evaluation also informs the school's planning and budgeting processes.
>
> **2.6** Evaluation is a simple, systematic, and sustained process that (a) identifies key educational and institutional outcomes (including learning outcomes for each degree program); (b) systematically and regularly gathers evidence related to each outcome (with a mixture of direct and indirect measures and quantitative and qualitative data); (c) engages appropriate stakeholders (especially faculty for educational outcomes) on a sustained basis to analyze and reflect upon how well the evidence indicates that each educational and institutional outcome is being achieved; and (d) uses those analyses and reflections for educational and institutional improvement.
>
> **2.7** Evaluation is formalized in one or more brief, cogently written plans that identify the parties responsible for evaluation and include a list of artifacts or instruments used to measure each outcome, a timeline that indicates how those artifacts or instruments are used, and clear benchmarks for evaluating success. Evaluation plans also indicate how often and by whom the evaluation plan is updated.
>
> **2.8** Evaluation is concerned with both educational quality and institutional effectiveness, though the primary focus for any theological school is on students— how well they are learning and how that learning helps them achieve appropriate personal and vocational goals. In the interests of public transparency, the school publishes a statement of educational effectiveness, giving evidence of educational quality by documenting appropriate areas of student achievement for each degree program (e.g., student learning outcomes data, graduation and placement rates, student satisfaction survey results) and by regularly updating that evidence with current information.

Even though the work of evaluation is covered most explicitly in Standard 2, words like "evaluate" and "evaluation" are found throughout the Standards. For example, Standard 1.2 discusses the relationship between mission and evaluation, Standard 3.8 talks about evaluating the curriculum, Standards 4-5 cover evaluation of specific degree programs, Standard 6.3 discusses evaluation of library services and resources, Standard 7.5 addresses evaluation of student services, Standard 8.5 mentions evaluation of faculty, Standard 9.5 covers evaluation of the governing board, and Standard 10.3 discusses the relationship between budgeting and evaluation. In addition, when we expand our view beyond the word evaluation, we see that the process of evaluation (especially as described in Standard 2.6, and unpacked in the next chapter of this guide) is indivisible from the Standards themselves, since the Standards are designed to "articulate principles of quality for graduate theological education that all schools meet in various ways" (from the *Preamble to the Standards*), and evaluation is the way that we know and can show both "how" and "how well" these principles are being met, in light of the mission and context of each school. Even when we aren't using the word "evaluation," we are necessarily doing the work of evaluation.

Speaking of language: as noted in the introduction to this Guide, the 2020 Standards of Accreditation intentionally do not separate "evaluation" (of the institution) from "assessment" (of student learning). This choice was made for two significant reasons. First, the membership noted that it desired one unified set of standards incorporating institutional and educational evaluation together, rather than something that separated institutional standards from educational and degree program standards as in the past. Schools experience this work to be interrelated, for their personnel and for their programming, and the design of the 2020 Standards reflects this. Second, and closely related, is that the membership recognizes that effective student learning cannot be separated from institutional effectiveness (and vice versa), and so it makes no sense to talk about one without also talking about the other. If the school isn't attending to its financial well-being, the learning experience almost always suffers; if the learning experience isn't adequate, the biggest recruiting budget in the world won't make a difference. Therefore, throughout the 2020 Standards of Accreditation (and, throughout this Guide), the word "evaluation" will refer both to institutional evaluation and to the evaluation of student learning, sometimes focusing on one more than the other and sometimes addressing the two equally. Schools may, of course, still use the word "assessment" to refer to the evaluation of student learning—or, really, may use any words that make sense in their context—since our focus is not on language but on practices and processes that serve our students and communities with our best efforts and aiming for our best outcomes.

In sum, throughout the Standards and throughout this Guide, the focus of evaluation 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? *Evaluation demonstrates to your stakeholders—including students and staff, donors and denominations—how well your school is living out its mission.*

Evaluation 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. Evaluation is how we know who we are, and how we can work to become closer to who we want to be.

The remainder of this guide will focus on how we do this work.

Chapter Two: Process of Evaluation ("How?")

As mentioned in the previous chapter, Standard 2.6 lays out the essential (and, expected) features of evaluation processes as follows:

> **2.6** Evaluation is a simple, systematic, and sustained process that (a) identifies key educational and institutional outcomes (including learning outcomes for each degree program); (b) systematically and regularly gathers evidence related to each outcome (with a mixture of direct and indirect measures and quantitative and qualitative data); (c) engages appropriate stakeholders (especially faculty for educational outcomes) on a sustained basis to analyze and reflect upon how well the evidence indicates that each educational and institutional outcome is being achieved; and (d) uses those analyses and reflections for educational and institutional improvement.

Another way to think of these four steps (a-d) is as *identification, information, interpretation, and implementation:*

a) *Identification* of desired goals or outcomes
b) *Information* on the achievement of those goals or outcomes
c) *Interpretation* of that information by key stakeholders
d) *Implementation* of those interpretations for meaningful improvement

In this chapter, we will look at each of these four elements in turn, and then conclude by reviewing a helpful (and, often, welcomed!) reminder about evaluation processes as a whole also articulated within Standard 2.6: evaluation should be simple, systematic, and sustained.

Step A. Identification of Desired Goals or Outcomes

As Standard 2.6 indicates, an evaluation process first "**identifies key educational and institutional outcomes (including learning outcomes for each degree program)**."

It perhaps goes without saying that good evaluation requires that we have first identified our goals; otherwise, how will we know whether we've arrived, or are even getting close? (Or, as Yogi Berra famously put it, "*If you don't know where you are going, you'll end up someplace else.*"). Yet a surprising number of schools, in evaluation summaries and self-study reports, are quick to laud their outcomes ("90% of our students can preach effectively!") without first identifying or prioritizing their goals. Being able to "preach effectively" might be a fantastic goal for a degree program focused on preparing future pastors, but perhaps less relevant for programs focused on lay leadership or counseling, and 90% might be great if earlier results were in the 70s or 80s, but perhaps not if your program aims for a higher number, or is focused on students who are already good preachers. Similarly, rushing to an institutional outcome ("We received $5,000 in alumni/ae donations this year" or "Our food service is consistently ranked best in our region") without articulating goals or outcomes is, perhaps, just a clanging

cymbal or noisy gong (Is $5,000 good or bad? Does the ranking of your food service matter to you?).

When it comes to institutional and educational goals or outcomes, particularly in relation to the Standards, you do not need to list everything that is important to you or everything that you hope your students will learn. Individual classes will still have their own course objectives; individual students or employees will have their own goals or aspirations; outside stakeholders (including denominations and employers) will have their own senses of what matters—all of which may or may not overlap with your broader institutional and educational goals. When we do the work of evaluation, we are seeking the articulation of goals and outcomes that focus on key factors that help demonstrate that you are doing what you say you are doing, including attention to the principles in the applicable Standards. In the context of educational evaluation, this means focusing on key factors that give a picture of a successful graduate for each degree program, in light of the expectations for that degree program in Standards 4 or 5.

It may be important at this point to highlight an important expectation in the ATS Commission Standards relative to evaluation of student learning. While student learning outcomes should focus on students, evaluation 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 evaluation 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 evaluating 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

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

those goals or outcomes helps the school achieve its mission. Put another way, evaluation *of individual students* tells us how well that student is doing; evaluation *of degree programs* tells us how well the school is doing. While the two are, again, obviously interrelated, we can't control all of the factors that impact an individual student's ability to thrive but can (and need to) address how well our degree and broader educational programs are serving those students.

Finally, when we articulate educational and institutional goals and outcomes, we do not want to set the bar so low in wording a goal or outcome that achievement of that goal or outcome is uninteresting or virtually automatic. The bar is best set where key constituents (particularly the faculty, for educational outcomes) believes it should be set, even if that means some results may be a bit discouraging at times. 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 goals or 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." Chapter Three will include more information about student learning outcomes in particular, and numerous resources on setting goals are available in the broader literature of higher education (with attention both to institutional and educational goals). From the Standards, we are reminded that goals should be relevant to the mission and context of each school (Standard 1.2); in practice, we know that outcomes often emerge from what we have learned from evaluation cycles in the past (i.e., where we have been helps us know where we seek to go next).

Step B. Information on the Achievement of those Goals or Outcomes

After goals and outcomes are identified, an evaluation process "**systematically and regularly gathers evidence related to each outcome (with a mixture of direct and indirect measures and quantitative and qualitative data)**" (Standard 2.6).

Collecting the right information is a critically important component of evaluation. We recall one seminary 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. Others note the persistence of something like a

> Collecting the right information is a critically important component of evaluation.

familiarity bias—we see or argue based on the evidence in front of us, rather than looking at all the angles (so I might believe that "all our students are happy" because the ones who interact with me seem happy, or that "we don't have bias on our campus" because I don't experience bias, or that "online learning isn't as good as onsite" because my past experience is only with onsite learning). A related challenge can be a *selection* bias, where we don't see the full picture because we don't look at the full picture (e.g., only asking happy students whether or not they're happy). *Causality* presents additional complications, as we may give ourselves credit ("90% of our students can preach effectively" must mean we have an amazing homiletics department) when the credit isn't actually ours (what if 95% were effective before they enrolled in our program?). Just as evaluation as a whole is an art more than a science, so too is the approach we take to the evidence we gather. In the end, the information or data collected—whether a statistic or a story, an intuition or a comprehensive investigation—must be sufficient to demonstrate how well the goal or outcome being assessed has been accomplished, remembering that our goal is to understand how well we are serving our students and our mission.

Put another way, if the first step in the evaluation process is to identify *where we want to go*, this next step is all about figuring out *how we would know whether we're going in the right direction*. Just as with setting goals, this means not just gathering whatever evidence happens to be around, or as much evidence as we can possibly collect, but rather on focusing in on what will help us know what we are doing and how well we are doing it. Standard 2.6 highlights two important pairs of data or kinds of information that are vital to understanding how well goals or outcomes are being achieved. Those two pairs are *qualitative and quantitative data* and *direct*

and indirect measures—and, among other things, these pairs remind us that a primary goal of this stage is to look from a variety of perspectives, so that the information we bring to the third step (interpretation) can be as broad and as helpful as possible.

Effective evaluation 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 (e.g., a rubric), or it might evaluate the effectiveness of its library based on outreach and usage statistics, but amplify this by adding a few vignettes of the experiences of different stakeholder groups. Too often it is assumed that for evaluation to be effective it must use only numbers, not narratives. In reality, there is a story behind every statistic, and some stories simply cannot be told well with statistics alone. Some schools also have the mistaken impression that if evaluation 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." To be sure, quantitative data can be very useful and should be part of any school's evaluation plan, but it is a bit of a fallacy to assume that because evaluation results are expressed numerically, they are not also "subjective." All results are subject to interpretation. In fact, without interpretation, evaluation results are fairly useless. Even the act of gathering quantitative (as well as qualitative) 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 evaluation 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 evaluation, see Chapter Three.

> 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 evaluation plans have both.

Effective evaluation also recognizes the interrelated nature of *direct and indirect measures* (both of which, by the way, can be expressed in either quantitative or qualitative ways). The simplest distinction between these two is this: direct measures assess *performance*, while indirect measures assess *perceptions*. For example, a juried music recital is a direct (and qualitative) measure of a student's performance, while a survey of students about the educational effectiveness of a music class is an indirect (and quantitative) measure of students' perceptions. A tally of the number of faculty publications is a direct (and quantitative) measure of scholarly service, while a vignette about how someone has been deeply touched by a faculty member's writings is an indirect (and qualitative) measure. Most ATS member schools tend to have a wealth of indirect measures of perceptions, the most common being course evaluations or other opinion surveys. However,

> *Direct measures* assess student *performance*; *indirect measures* assess student *perceptions*.

evaluation that is based only on indirect measures runs the risk of not being able to document *what* (or *that*) goals were met, only what someone *thinks* about those goals being met. On the other hand, indirect measures do help a school understand the level of stakeholder satisfaction that exists—and in this consumer-oriented, higher education landscape a school not concerned about stakeholder satisfaction runs the risk of losing students and employees. In the area of educational evaluation, 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 or not, even if they ought not be taken as the only judge—especially since students (like all of us) tend to both overestimate and underestimate our own impact and accomplishments.

Of course, quantitative vs. qualitative and direct vs. indirect are not the only facets we bring to the work of data collection. We want to look from as many angles as needed to see the full scope of our work, and diversity of perspectives (and disaggregation of experiences) helps us build a better data picture. Remember that the desire here is not just to "gather opinions" or "make a case" but rather to explore, with curiosity and openness, in ways that help you learn, understand, and (eventually) act in response to the outcomes and goals you have articulated. As part of this, schools may find it very helpful to distinguish evaluation data by gender, age, race/ethnicity, or other demographic information. For example, such data can prove very useful in devising strategies to improve student learning, since needs or experiences might vary significantly between demographic groups (e.g., "day students" and "night students" might have very different experiences, as might "lay students" vs "clergy" or "first career" vs "second career"). Some schools may also find it very helpful to distinguish learning evaluation data in terms of delivery modalities (onsite, offsite, and online), since again the needs or experiences of students in various settings may be different. Having that kind of information available to key stakeholders (including faculty) can assist in the interpretation process—the subject of the next part of this guide.

Regarding the overall collection of evaluation 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 that manage thousands of pieces of information. Here, too, the watchwords should be simple, sustainable, and sufficient. Similarly, schools (especially those working toward a self-study report) often ask how many years' worth of data should be included in the report. While that varies due to several factors, there needs to be enough data to show what the trends are—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 evaluation isn't for evaluation's sake, but rather for ongoing educational and institutional improvement as you seek to live out your mission.

Step C. Interpretation of that Information by Key Stakeholders

After goals are identified and information is gathered, an evaluation process "**engages appropriate stakeholders (especially faculty for educational outcomes) on a sustained basis to analyze and reflect upon how well the evidence indicates that each educational and institutional outcome is being achieved**" (Standard 2.6).

In order to be successful with this step, schools must first identify, and then engage, key stakeholders. In the area of educational evaluation, this means the faculty—as a whole. Too many reports on the evaluation of student learning reviewed by the ATS Board of Commissioners "fail the faculty test" because there is no evidence that faculty are appropriately and seriously engaged in the evaluation process. The ATS Commission Standards place such a high priority on faculty involvement in evaluation

> The ATS Commission Standards place a high priority on faculty involvement in evaluation because they place a high priority on faculty in student learning.

because they place such a high priority on faculty in the student learning process. While some aspects of evaluation can be delegated (primarily data collection), the interpretation of evaluation information rests best with the faculty who know best the aims of theological education. Faculty also have the best understanding of what students should be learning, and of what might be affecting that learning. Evaluation 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 non-representative sample ("bad classes" do happen) or was the result of a one-time experiment (since abandoned) or was due to a fault of the evaluation 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. A final reason to give faculty such significant responsibility in interpreting data about educational effectiveness is that the faculty are the ones who will need to make any resulting changes to improve student learning.

> Evaluation data can lead to poor decisions if not interpreted correctly and contextually.

While administrators or trustees or other stakeholders 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.

A second essential group of stakeholders, for both educational evaluation and institutional evaluation, is the school's governing body. Whether through the school's trustees as a whole or through an appropriate subcommittee (e.g., a seminary committee or an academic affairs committee), Standard 9.5 states that the governing body "ensures that the school's mission and educational and institutional

> The governing board of the school "ensures that the school's mission and educational and institutional outcomes are regularly evaluated and that the results are used to better achieve the school's mission and improve its various outcomes" (Standard 9.5).

outcomes are regularly evaluated and that the results are used to better achieve the school's mission and improve its various outcomes." For institutional evaluation, this work would likely be done in partnership with the school's administration, under the umbrella of shared governance (Standards 9.6-7). For educational evaluation, the partnership would be with the school's faculty (Standard 9.8). 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 results of the evaluation are interpreted. If trustees represent a school's various constituencies, especially churches or employers, 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 evaluation 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.

Even as faculty and trustees serve as key stakeholders for the work of evaluation (and, particularly, for the review and analysis of the information gained from evaluation), they are not the only ones. For example, Standard 3.7 expects that "the school demonstrates an intentionally collaborative approach to student learning and formation by developing a cohesive and holistic curriculum, regardless of modality, that involves faculty and, as appropriate to the school's context and degree programs, librarians, student services personnel, field educators, and others—both in designing and in evaluating the curriculum." This focus on the school's context is important and significant. At some schools, this sort of involvement may take the form of "all hands on deck," whereas at others it may be more structured via layers of committees or processes of representation. At some schools, the "and others" might include students, alumni/ae, donors, and/or community members; at others, the work may be more narrowly focused. Again, this is where it is helpful to remind ourselves of the purpose of evaluation—"to discern how well the various aspects of the school's mission are being achieved and how its educational and institutional outcomes could be maintained if met or improved if not met" (Standard 2.5)—which then allows us to consider who needs to be part of this work, and in what ways, in order for us to best achieve our purpose.

Once the stakeholders have been identified, they can turn to the work of analyzing and reflecting upon how well the evidence indicates that each educational and institutional outcome is being achieved (Standard 2.6). Here it is essential to remember that data does not speak for itself. Every data point and its connection to another can tell multiple stories, and so curiosity and diversity of perspective are paramount in this interpretive phase. Ask hard questions of the data, particularly with your evaluation goals in mind (what are you hoping to see; what are you hoping to say). For example, a school might lean into someone with expertise in data collection and analysis (e.g., an administrator, staff person, faculty member, or department) who coordinates the overall evaluation process, especially the collection of data, and then provides it to key stakeholders in a synthesized, manageable form—not for them to

simply sign off on it, but to wrestle collectively with what the data means (particularly in light of the mission and context of the school and the stated goals/outcomes of each program) and, from that wrestling, to then discern next steps. Many ATS member schools have found it helpful to designate an "assessment day" or "evaluation retreat" each year, so that they can spend a day or two of focused time together interpreting artifacts, evaluating data, and planning actions (the topic of the next section of this guide). Others build moments of reflection and evaluation into each regular meeting (e.g., faculty, staff, trustees), or add a "what we're learning" corner in newsletters or internal intranet spaces, or regularly schedule debriefing and evaluation meetings after key events and initiatives. This is another place where evaluation is not "one size fits all"–each school needs to identify and enact a process of analyzing and reflecting on data in a way that fits your mission and context and helps you learn and understand how well your goals are being met, and to imagine what you might do to better reach the desired outcomes.

Step D. Implementation of those Interpretations for Meaningful Improvement

Finally, even once goals are identified, information is gathered, and evidence is interpreted, the evaluation process isn't complete until the school "**uses those analyses and reflections for educational and institutional improvement**" (Standard 2.6).

When all is said and done about evaluation, not answering well the "So what?" question makes the entire evaluation process devolve into much ado about nothing. If evaluation does not lead to improvement, the process seems pointless. Otherwise, the common critique that evaluation is just about satisfying some accreditation regulation is a valid one. Remember that accreditors, however, are not looking 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 evaluation efforts is this: what difference has this made for your students and school? They will often ask administrators, faculty, staff, and trustees to cite a few examples of how evaluation led to improvement in student learning and/or in departmental or institutional effectiveness. Regardless of how elaborate the process and extensive the paperwork, if no one can readily list key changes that came from evaluation, then that evaluation is not effective.

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

To be sure, often the outcome of a review and analysis of evaluation data 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 focus should be on how these aspects of the school's educational and institutional outcomes can be maintained (per Standard 2.5). It should be noted that "maintained" is not the same as "doing nothing." If you are highly effective in one area because of skilled employees in that department, the outcome of the evaluation process might be to consider how to support and retain those employees, or how to recruit (or home-grow) employees with similar skills if the current employees retire or leave. If a degree program is highly effective in one modality or with one segment of the student population, the question might become how to translate this effectiveness into other modalities or other students, or even to other degree programs or non-degree opportunities. Learning more about your

successes can also help you tell your story more persuasively (with data and details), which could also help with student or donor recruitment. And, of course, every good evaluation result can be a cause for praise and is worth celebrating. Discovering what we are doing well is just as important—and, just as interesting—as discovering what is not living up to expectations.

It goes without saying, though, that not all of the results from our evaluation processes will feel like "good" ones. Sometimes this is because the evaluation process itself needs revision; perhaps the goal or outcome isn't clear or appropriate, or the artifacts or evaluation process aren't connecting well to those goals, or the interpretation process isn't happening effectively. An evaluation process—just like everything else—needs to be evaluated and improved as needed. Other times, the results will be disappointing because the effectiveness of the school, in one way or another, isn't living up to expectations. The ATS Board of Commissioners 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 make data-driven decisions and implement those changes that are truly needed—rather than either pretending that everything is "fine" or focusing efforts on tasks that aren't likely to make a difference—which then helps us better live into our missions.

We repeat here the point made earlier: effective evaluation 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 evaluation 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

> Perhaps the most fundamental question to ask is whether your evaluation helps your school achieve your mission.

depends on *your* mission. In short, evaluation is still always about improvement. The first three components in the evaluation cycle mean very little if schools don't "close the loop" by using the information to implement meaningful changes. Perhaps the most fundamental question to ask is whether your evaluation process helps your school achieve your mission. That is why the ATS Board of Commissioners pays such close attention to member schools' evaluation efforts. Evaluation is about mission.

Don't Forget: Simple, Systematic, and Sustained
Standard 2.6 also includes three key characteristics of effective evaluation: it is "a simple, systematic, and sustained process...." This means we should be asking these three questions throughout all four steps of the evaluation process.
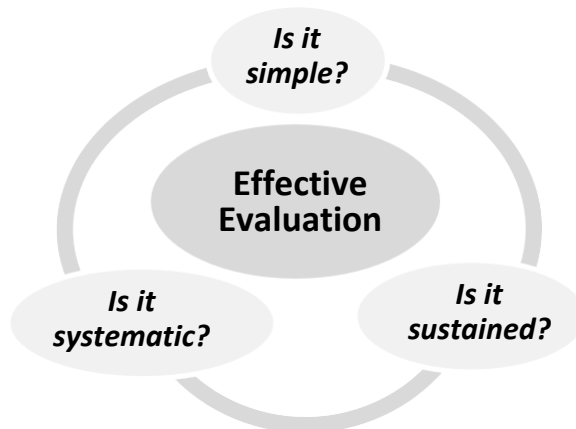
Is it simple? The ATS Board of Commissioners gives no extra credit for extra length. In fact, some of the *least* effective evaluation plans it has seen are some of the *longest*. It is just not reasonable nor necessary for most schools to have elaborate plans or results that run into dozens, scores, or even hundreds of pages. Even for larger schools, it is important that clarity receive priority over length. Some of the most effective evaluation plans that the ATS Board of Commissioners have seen are only a few pages long (though the raw data behind those summaries might be more extensive). In part, this is a pragmatic concern: most key stakeholders (including faculty and trustees) simply do not have the time to pore over pages upon pages (or screen after screen) of evaluation data. And if these stakeholders are not seriously engaged in evaluation with meaningful conversations about the results, then the plan is pointless and the results useless.

A simple evaluation plan often includes focusing in on a manageable number of goals and outcomes (see more on Student Learning Outcomes in Chapter Three). It may also include a rotation schedule, so that stakeholders are not reviewing every goal and outcome every time they engage in evaluation (one good practice may be to review a "dashboard" of all items regularly but to go into detail with only one or two outcomes at a time). It is also helpful to focus on simplicity in data collection: most schools already have a wealth of ongoing evaluation information but may not realize it. For example, a senior reflection paper or a case study at an internship site might provide very useful data about a particular student learning outcome, but this artifact might be known only to the faculty member who assigns it; a denomination might already gather placement and longevity data so that you don't need to collect it yourself; ATS might have instruments or data that saves you the effort of creating something from scratch. Before collecting new data, it is helpful to pause and consider what existing data and resources you have (or have access to), as well as whether there are existing ways to collect whatever new data you need. Similarly, at the evaluation stage, you may be able to draw on the work of existing structures (e.g., a curriculum committee or student retention committee may be able to take the lead in some evaluative work related to their areas of responsibility) or rhythms of the calendar (e.g., a faculty retreat might be an opportunity for an add-on session on evaluation of student learning). Another option for the work of interpretation would be a "divide and conquer" (or, more accurately, "divide and share") strategy so that the work is distributed in manageable amounts. Again, as always, all of the stages of the evaluation loop should be

appropriate to the mission and context of your school—as complex as necessary, but not more complex than necessary.

Is it systematic? Neither the evaluation process as a whole, nor any of its constitutive parts, are intended to be haphazard or arbitrary. Not only do the Standards talk about the evaluation process—and data collection in particular—as being necessarily systematic (Standard 2.6), but when we return to the overall goal of evaluation ("to discern how well the various aspects of the school's mission are being achieved and how its educational and institutional outcomes could be maintained if met or improved if not met" per Standard 2.5), it is clear that we must look broadly at what we're doing if we're going to look effectively at how we're doing, particularly if our mission is at stake. Too many evaluation plans look only at what is easiest to track (as discussed earlier) or at what seems most important or urgent (a new or flagship degree program, a financial stressor, a troublesome employee) rather than at the embodiment of our mission more broadly (the bread-and-butter programs, the stable-but-essential personnel, the marks of formation most desired by our faith community). If we're going to ask whether a degree program is effective or whether our institutional goals are being met, we need to take a systematic look at what we are doing, rather than just chasing the shiny object.

In addition, the value of being "systematic" also applies to how we articulate and keep track of our work. Schools are often criticized for not having a *culture of evaluation*, when more often

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

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. Appendix C also provides more common examples of how schools might document evaluation results.

Is it sustained? Closely related to the question of whether an evaluation program is systemic is whether it is (and, can be) sustained over time. A key question is not only "How are you doing?" but also "Where have you been and where are you going?" While a first essential step

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

is for schools to have an evaluation plan, accreditors and other stakeholders long ago stopped being satisfied with simply a *plan* for doing evaluation. It is evaluation *practices and results* over time that matter, especially how schools use those results over time to make institutional and educational improvements for its future, and how they gather and review data *after* making improvements to ensure that those interventions had the desired outcomes. One of the more problematic (as well as least effective) evaluation practices that visiting committees sometimes discover is a school that has gone through an intensive evaluation process to prepare for an

accreditation 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. 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 evaluation than it is to do one intense attempt every now-and-again.

In preparing evaluation plans and strategies, then, schools will want to think about what sustainability looks like in their institutional context, taking into account the workload and regular rhythms of their faculty and staff, among many other factors. Schools will want to gather and evaluate data often enough that they can clearly demonstrate program effectiveness and can quickly assess the effect that any changes (including faculty changes, course revisions, new textbooks, a shifting financial environment, and so on) are having on the institution, including on what students are learning. When significant changes are happening, schools may wish to "speed up" their evaluation cycles (e.g., reviewing a new degree program annually through the first few cohorts of students, keeping close tabs on employee wellness and student satisfaction during financial exigency). But, committing to evaluation as a sustained practice means doing your "annual checkup" even when you're feeling well. If nothing else, these regular practices provide a baseline later on when a point of "healthy comparison" might be needed, in addition to attending to the ways in which we can always seek improvement. And, more often, regular practices of evaluation can help us catch changes and challenges before they become more significant and can help us focus our efforts on continuous improvement in ways that are simple, systematic, and sustained.

Chapter Three: Practices of Evaluation ("What?")

As noted earlier, the 2020 Commission Standards of Accreditation focus on principles more than practices, inviting schools to develop practices that meet those principles in ways that are contextually appropriate and attend to the mission of the school and the needs of its stakeholders. This chapter will discuss some of the practices that schools might consider as they engage in evaluation. Most of these focus on the evaluation of student learning (what the earlier versions of the Standards referred to as "assessment"), but some of these practices (e.g., benchmarks) are also useful in other sorts of evaluative work. Some of these practices are ubiquitous, in that it is hard to imagine an effective evaluation process without them; some (like a statement of educational effectiveness) are required by the Standards and/or by the Commission Policies. As with other elements of this Reflective Guide, schools are reminded that the ATS Commission Standards on Accrediting, not this guide, constitute the final word on evaluation for member schools—and, as noted earlier, there are many good ways to engage evaluation, and not all of them can or will be included here, and not all of the sections that follow will be relevant or contextually appropriate for all schools.

Student Learning Outcomes

Throughout the literature on evaluation, two related terms are frequently used: *degree program goals* and *student learning outcomes*. While these two terms are sometimes used distinctively (e.g., student learning outcomes might include both degree program goals and institutional learning goals or denominational learning goals), they are more often used synonymously, and we encourage schools to not get stuck on vocabulary. Regardless of the terms used, the evaluation 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.). 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 (for example, 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 sections of Standard 4 (for masters degrees or Standard 5 (for doctoral degrees).

Schools may also have institutional goals related to their educational program that are not directly related to student learning. Some might be in the category of "student outcomes" (as differentiated from "student *learning* outcomes"), such as graduation and placement rates, or the cost and time to complete a degree, including issues of student debt—these are goals that aren't directly related to learning but that impact the student experience. Others might be in the category of "program goals" and might focus on non-public, internal goals (e.g., financial results or admissions targets) or might be broader (e.g., that the program embodies diversity and that alums find placements in prestigious settings). These 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.

Student learning outcomes tend to focus on what students actually learn through their curricular and co-curricular experiences. They typically require 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." Here 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 Standards require each degree program have "clearly articulated learning outcomes" (see Standards 4.3, 4.8, 4.11, 5.3, 5.9, and 5.13). 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 evaluation strategies focused on "those ways."

To keep evaluation 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 and other expectations described in the applicable sections of Standards 3-5. 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. To keep things simple, some schools may design degree programs whose goals or outcomes overlap. While some overlap is acceptable (and even unavoidable), it is not acceptable simply to use the same exact set of degree program goals or student learning outcomes for more than one degree—since the degrees are different, the outcomes should be different. 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. In all cases, student learning outcomes 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?

Direct Measures of Student Learning

As noted in chapter two, evaluation should include both direct and indirect measures (and, either can be expressed in quantitative or qualitative ways). In the context of student learning, the simplest distinction between these two is this: direct measures assess *student performance*, while indirect measures assess *student perceptions*. Both are important, and both are required by Standard 2.6. But schools sometimes struggle with finding appropriate direct measures. 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.

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 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 include 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

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

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, etc. Those results can be 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—and then, this data can be supplemented with additional instruments or data sources that attend to other degree learning goals in ways that attend to the distinctive mission of each school.

Grading and Evaluation

We still occasionally encounter schools that try to use course grades as part of an evaluation strategy; some will even argue that it could be considered a direct measure. To be sure, *course grades are a form of assessment*, and they should not be simply dismissed as of no value. But course grades are not sufficient to tell the whole story of what students are learning, and they

are rarely useful as part of a systematic evaluation process. Five reasons for this come to mind: (1) course grades can be rather arbitrary (just ask any student); (2) course grades typically measure more than just student learning (e.g., attendance, timeliness, course participation); (3) course grades usually signify the professional perspective of an individual instructor rather than the evaluative consensus of faculty as a whole, (4) course grades are notorious for inflation ("Grades here run the gamut from A- to A+"), and (5) course grades focus on 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 co-curricular experiences and the cumulative effect of a student's learning) if these are relied on as part of a systematic evaluation process.

Course-Embedded Assessments

However, while course grades fail us, it is possible to use course-embedded assessments to gather meaningful evaluation data. This 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, you can draw on those student assignments as part of your evaluation work. You don't need to develop an entirely new artifact or evaluation tool—rather, you can reuse that assignment (e.g., anonymize the responses and have a faculty jury review it with the degree learning outcomes in mind) or you draw data from it (e.g., have the instructor fill out a rubric about how well the student demonstrated the learning outcome, separated from any grade adjustments due to other factors such as timeliness or sentence construction). This same principle, of reusing or drawing data from existing sources, can be found in other settings as well. For example, students might demonstrate achievement of learning outcomes when they participate in the co-curricular life of the school (leading chapel, organizing outreach activities, participating in student government). 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, or as part of recommending students for ordination or graduation; the work of evaluation can consider ways to gather and analyze this data in regular and ongoing ways.

Portfolios

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. 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 also provide the opportunity for faculty to reflect both on *formative* (in-process) assessment and *summative* (end-product) assessment, for example by noting how students improve from their first year of study to graduation. Portfolios can serve multiple functions: for the student, it becomes a record of their work throughout the program (which might be useful in future settings, such as job searches or ordination processes, or might just serve as a source of celebration and accomplishment for the

student); for the advisor or faculty, it can serve as a reminder of the individual student's successes and challenges (which can be useful when recommending that student for graduation, ordination, or employment); for the evaluation process, it can be anonymized and aggregated to see how well all students are achieving degree program goals or outcomes, not just how well each individual student is achieving them. In particular, a portfolio system can be quite helpful for educational evaluation when specific assignments are linked to specific learning goals (e.g., exegesis papers that demonstrate students' ability 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).

The difficulty with portfolios is that they can become logistically complicated (even with good software packages and highly motivated students, schools sometimes report spending more time on the *administration* of a portfolio system than on the *evaluation* of the portfolios) and the artifacts that are collected need to be closely connected to degree program goals or student learning outcomes in order to be useful. 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 (something faculty often report as frustrating and bordering on busywork). In addition, depending on the institutional context, it may be quite challenging to use portfolios for degree program evaluation if students have taken different courses from each other, or submitted different assignments, or moved through the degree program at different speeds or with transfer credits. Because of these challenges, portfolios often do not contribute to an evaluation process that is "simple, systematic, and sustained." Schools considering implementing a portfolio system might find it helpful to consult with peer schools to learn from their successes and challenges or might begin with a small pilot group of students or with just a single degree program before rolling it out more fully.

Curricular Map

One strategy in linking educational evaluation to mission is to develop a curricular map. 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 evaluation 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 professionally-oriented degree programs—are best delivered by experiences outside the classroom, such as supervised ministry, CPE, or formational experiences. This also means that a curricular map should be developed not by core faculty alone but rather in conversation with administrators and other staff, particularly those who oversee or participate in those non-classroom experiences, as well as with other stakeholders as appropriate.
.

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. 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, or that a new goal/outcome should be added, or it might be that the course meets a need other than a degree program goal. 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 could either mean needless redundancy or necessary reinforcement. To address such possibilities, some faculty take the curricular map a step further and instead of using simple "X's" in boxes use a more informative code, such as "I" (this course introduces that goal/outcome), "R" (reinforces that goal/outcome), or "C" (is the culmination of that goal/outcome in this program). And, above all, it is important to remember that the goal of the curricular map is not to do it "right" or to "get the best score" but rather to generate conversation about student learning, for the sake of improving student learning, in light of the mission and context of the school.

> 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.

Rubrics

Rubrics are simply guidelines for rating student performance. Some rubrics are used for specific course-embedded assessments (e.g., exegesis paper) in order to bring greater clarity and consistency to evaluation. 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 evaluation), those involved in grading that exegetical project develop a clear grading rubric that clarifies what is expected of students to achieve a B, or an A, or what would constitute a C or less. 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 evaluation. Such rubrics can also save faculty time because they are not writing the same comment over and over again on dozens of different assignments.

Other rubrics are used for evaluating 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). 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 outcome. As noted earlier, course grades usually signify the professional perspective of an individual instructor rather than the evaluative consensus of faculty as a whole. Rubrics that are developed by the faculty as a whole—or, by a group of faculty (in a particular department, for example)—can bring a broader perspective, which is essential for degree program evaluation.

It is worth noting at this point that rubrics, like evaluation, 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 evaluation, rubrics for a single course-embedded assessment 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 evaluation 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.

Benchmarks

In using both quantitative and qualitative data through both direct and indirect measures, it is important for effective evaluation to have some sense not only of what information is needed, but what "criteria for success" will be used to evaluate or interpret that information. As Standard 2.7 notes, each evaluation plan (including each degree program goal or learning outcome) needs some kind of "performance indicator" or "benchmark" so that evaluation results can be interpreted meaningfully.

> It is important for effective evaluation to have some sense not only of what information is needed, but what "criteria for success" will be used to evaluate or interpret that information meaningfully.

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% of the group will score at least a 3.5)?

Benchmarks simply offer us a way to begin interpreting the data we have gathered, with a sense of what improvement (if any) may be needed.

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. For example, a performance indicator or benchmark for the previously cited goal or outcome might be this: "At least 85% 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." As this example shows, benchmarks can draw on quantitative and/or qualitative data and indirect and/or direct measures and can put multiple pieces of data in conversation with each other.

As noted earlier with goals or outcomes, performance levels should be set at reasonable rates. For example, to set the levels at 100% 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 program is ineffective. On the other hand, setting the level at, say, only 50% 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.

Quantitative Data

In chapter two, we discuss the use of quantitative and qualitative data. It may be helpful here to say a few more words about numbers—at least these four.

First, schools often wonder how big a number they need to have adequate evaluation results. Some might argue, for example, that since they only graduate a handful of students each year, then that number is too small to yield meaningful evaluation results. In reality, small numbers mean a school can evaluate the entire population and not have to rely on a sample. Evaluating what all five of your graduates have learned is just as meaningful, if not more so, than evaluating what twenty five of thirty graduates have learned. However, small numbers can also present a challenge for some schools. For example, a school evaluating 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 evaluate every student every year in every area. Smaller samples of larger numbers can be a very efficient and equally effective way to conduct evaluation—as long as you make sure that the sampling is legitimate. For example, a school might choose to have faculty review ten out of thirty student assignments as an indicator of how that assignment achieved a specific student learning outcome for a specific degree program. Evaluating just ten out of thirty (or five out of twelve; there is no required ratio) is perfectly legitimate as long as those samples are not inappropriately chosen, such as choosing assignments only from top 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, so that confidentiality can be appropriately preserved.

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% of our 40 graduates last year indicated… Or, our faculty reviewed ten out of thirty 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% 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 evaluation 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.

Statement of Educational Effectiveness

One required practice for all schools is the public statement of educational effectiveness. As noted by Standard 2.8, "In the interests of public transparency, the school publishes a statement of educational effectiveness, giving evidence of educational quality by documenting appropriate areas of student

> The principle behind this standard is that evaluation is not only about improvement, but also about accountability.

achievement for each degree program (e.g., student learning outcomes data, graduation and placement rates, student satisfaction survey results) and by regularly updating that evidence

with current information." The principle behind this standard is that evaluation is not only about improvement, but also about accountability.

As indicated in the [Self-Study Ideas](#) for Standard 2.8:
> A strong public statement of educational effectiveness includes some or all of the following information for each degree program: data from the school's evaluation of degree learning outcomes, ratings from student satisfaction surveys and/or exit interviews, alumni/ae and/or employer feedback, graduation rates, placement rates, time to completion rates, and other relevant data regarding educational effectiveness. Information from the ATS Graduating Student Questionnaire and Alumni/ae Questionnaire could also be helpful, as well as student and alumni/ae testimonies. The public statement of educational effectiveness should be available in an easily accessible location (e.g., in a prominent location on the school's public website) and should be reviewed and updated regularly (see *Policies and Procedures*, VII.B.1). See Standard 7.11 for a discussion of placement rate data.

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. 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% 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.

Many schools choose to make public in their statement 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. Some combinations of all these can make a compelling case for educational effectiveness and help the school "tell its story well."

Additional ATS Resources
As mentioned in the introduction, ATS makes available a variety of resources to help schools with the important work of evaluation. In addition to the helpful role your Commission staff liaison can play, we also want to be sure to mention the available instruments and tools that can help you with data collection and interpretation. The ATS student questionnaires (Entering Student Questionnaire, Graduating Student Questionnaire, and Alumni/ae Questionnaire) are

one example. While many of the questions focus on indirect measures of student satisfaction, some direct measures are also included (e.g., regarding work history, educational debt, and placement). The use of these instruments by ATS member schools is strictly voluntary. For schools that do choose to use the ATS *Q's*, 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—Questionnaires—Resources: Interpreting the Questionnaires. Archived report results for the *Q's* from all ATS member schools who choose to use those instruments can also be found under Resources—Student Data .

Schools working on institutional evaluation will likely find the Strategic Information Report (SIR) to be especially helpful. The SIR is designed to provide ATS member schools with a variety of strategic indicators to help assess their overall financial strengths and performances. The SIR can also be used as a tool in your school's strategic planning process by providing ten years of institutional data along with benchmark comparatives to other theological schools. The Institutional Peer Profile Report (IPPR) provides comparisons of annual data reported by ATS schools with those of other ATS schools that are identified by the requesting institution as peer institutions. And the ATS data visualization tool allows users to slice and view ATS enrollment data in a variety of ways. Through this tool, ATS has made information already publicly available from the annual data tables now accessible in one location and across years. Users have access to three types of visualizations and a variety of slicers, such as primary denominational family, country, and ecclesial family of the school, as well as size and Carnegie classification. Data can also be visualized by a few characteristics of the student, such as race and gender.

Finally, the Association maintains a webpage with tools and resources for Evaluation and Assessment Coordinators. Among these, we would highlight the Assessment Workbook for Roman Catholic Seminaries, which explores educational evaluation with attention to both the ATS Commission Standards and also the United States Conference of Catholic Bishops' Program of Priestly Formation (PPF) and Canadian Conference of Catholic Bishops' Program for Priestly Formation (CPPF). This text is accompanied by a selection of assessment case studies. Because the workbook and case studies focus on student formation broadly (including human, spiritual, intellectual, and pastoral elements), it can be useful to a wide range of institutions, including Catholic schools that offer programs other than priestly formation, as well as non-Catholic institutions as they try to imagine new ways to assess spiritual formation or personal character. The site also includes a bibliography of additional resources, such as those provided by the Center for Teaching and Learning at Indiana University-Purdue University Indianapolis and the National Institute for Learning Outcomes Assessment (NILOA), and information about the Assessment Coordinator's conversation space on Engage ATS, an online interactive community for those who serve at ATS member schools where you can ask questions and share learnings with your peers.

Final Thoughts

As we conclude, we return again to the line from *The Dead Poets Society*: "Poetry…reminds us that the powerful play goes on, and you may contribute a verse." Evaluation is not numbers on a Likert scale. Evaluation 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 evaluation is important. What will your verse be?

Appendices

Appendix A: Checklist for Effective Assessment

The following checklist for effective assessment was developed for the previous version of the Standards (2012), to help schools quickly review their process of educational evaluation ("assessment"). While the Standards have changed, this checklist still summarizes and synthesizes good practices, based on the process of evaluation described in the new Standards and throughout this guide. This checklist can never be a substitute for the thoughtful and purposeful work that effective evaluation requires, but it can be a way of remembering key components of effective evaluation (particularly of student learning). Schools may find it helpful to use this as a starting point for further conversation, reflection, and action.

Checklist for Effective Assessment			
1. Effective assessment plan and process	**YES**	**NO**	**IN PROCESS**
a) Appropriate mandate/oversight from governing board?			
b) Designated person or office to coordinate assessment?			
c) Plan/process reflects school's mission, ethos, resources, and size?			
d) Faculty play central role in assessment process?			
e) Is the plan/process: simple, sustainable, sufficient/adequate?			
f) Is the plan/process itself assessed and changed as needed?			
2. Identification of desired goals/outcomes—for each degree			
a) Goals/outcomes clearly stated for each degree program?			
b) Goals/outcomes reflect appropriate Standards?			
c) Goals/outcomes are expressed in demonstrable terms?			
d) Goals/outcomes are manageable in number?			
e) Goals/outcomes are decided by faculty in those programs?			
3. Information on the achievement of those goals/outcomes			
a) Information comes from both direct and indirect measures?			
b) Information includes both quantitative and qualitative data?			
c) Information uses some form of "benchmarks"?			
d) Information includes data about completions/placement?			
e) Information does not rely mostly on course grades, uses rubrics?			
f) Information about individual students is kept confidential?			
4. Interpretation of that information by key players, especially faculty			
a) The "key players," especially faculty, are engaged in assessment?			
b) Faculty have substantive conversations about assessment data?			
c) Faculty review regularly the implications of this information?			
d) Faculty interpret raw data in appropriate ways in light of mission?			
5. Implementation of those interpretations for meaningful improvement			
a) Faculty interpretations lead to improvements in student learning?			
b) School provides clear examples of such improvements?			
c) Board ensures results shape educational/institutional decisions?			
d) Assessment plan/process and results are clearly documented?			
e) School summarizes key assessment results publicly and regularly?			

Appendix B: Sample Summaries of Evaluation 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 evaluation 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.

Many schools find the classic "Nichols model," or some variation of his "5-column model," to be a helpful way to maintain a *simple* evaluation process. James O. Nichols, a university director of institutional effectiveness, pioneered in the early 1990s a five-step process for connecting the evaluation 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.

Sample #1: Assessment Results Summary for an MDiv
Chart Format Using Nichols Model

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 4.3)
	Direct Measures:
	  Entering Bible exam
	  Exit Bible exam
	  Exegesis Paper in BL702
	  Juried Review Sr. Sermon
	Indirect Measures:
	  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 4.3)
	Direct Measures:
	  Global engagement project
	
	  Cultural awareness scale
	
	Indirect Measures:
	  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.	*Direct Measures:*
	  Field mentor evaluation
	
	  Faculty advisor evaluation
	
	Indirect Measures:
	  Exit survey/interview
		Mentor rating of "meets or exceeds expectations"
		Faculty rating of "meets or exceeds expectations"
		
		Average rating of 4 of 5
			No mentor ratings were collected
			Faculty ratings met that benchmark
			
			Averaged 3.9 of 5
				Added mentor training
				
				No changes; met
				
				Changed benchmark
				  (grads too "humble")

(Standard 4.3)	Annual Alumni Survey	Average rating of 4 of 5	Averaged 4.1 of 5	No changes; met (alum noted "humility")
	ATS GSQ Table 19	Average rating of 4 of 5	Averaged 3.9 of 5	Appointed faculty taskforce to review
4. Graduates will…	[note this outcome deleted for brevity sake from this example]			

* The assessment results listed here are simplified samples. An actual assessment report should specify the time period covered (e.g., 2021 graduates), as well as give some indication of the total population (e.g., 15 total graduates from 2021) 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).

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:
a) Direct Measure: Exegesis paper from MDiv portfolio.
b) Direct Measure: Capstone research project, scored using faculty-wide rubric developed last year.
c) 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.
d) 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).
e) Indirect Measure: Exit interviews from MDiv graduates.

Criteria for Success: [NOTE: "Criteria for Success" = "Performance Indicator" or "Benchmark"]
a) 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.
b) 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.
c) 80% of graduating students taking the GSQ will report a score of 4.0 or above (out of 5)
d) 75% 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 ten years.
e) A random sample of exit interviews will generally be positive, with affirmations and helpful suggestions far outweighing concerns or complaints.

Assessment Results:
a) MDiv grads averaged above "acceptable" on rubric for this artifact in the MDiv portfolio, with 30% scored at the "excellent" level.
b) MDiv grads averaged 3.5 on capstone rubric; 1 of 20 grads rated 1.0 (unacceptable), with that one student viewed as an "outlier."
c) 95% 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).
d) 73% 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% over the last five years.
e) 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:
a) Faculty discussed this at length and concluded that this is a strong area for our MDiv grads; they also recommended that the rubric (introduced 3 years ago) be revised with clearer criteria.

b) 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.

c) The 95% 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.

d) 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 a fifth of the students (22%) merits further discussion. A faculty task force was appointed by the dean last fall to bring back recommendations.

e) 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 Standard 4.3.