part one

the fundamental goals of review and evaluation

1

the accountability culture in higher education

For much of the history of higher education, conducting student or faculty surveys—even conducting formal student examinations—was relatively rare. In some systems, course examinations either didn't occur at all or didn't play a major role in a student's progress; the important test was the set of comprehensive examinations that occurred just before the granting of a degree. (See, for example, Amano, 1990; Kehm, 2001; Wiseman, 1961; Min and Xiuwen, 2001.) Even well into the twentieth century at American universities, many courses based a student's grade solely on a final examination or research paper (Smallwood, 1935). Some professors also gave a midterm exam, and some courses included quizzes or tests after each major unit, but the notion of frequent grading opportunities was not particularly common at most colleges and universities.

Even less common were student satisfaction surveys, forms evaluating professors or administrators, and structured performance reviews of staff members. The assumption was that the faculty knew far better than the students what needed to be taught in university-level courses, and administrators were responsible for making sure that the faculty taught those courses effectively. They may not have

evaluated those faculty members in a regular and former manner, but if they gained some sense that a particular instructor was ineffective, they responded with either advice or termination of the person's contract. Systematic efforts to assess the effectiveness of academic programs or evaluate the continual improvement in a faculty member's teaching, research, and service were all but unknown. And then, seemingly all at once, everything seemed to change.

○ why everyone always seems to be evaluating everyone else

If you enter a university in many parts of the world today, you're likely to encounter a bewildering array of surveys, assessment instruments, and examinations. Whatever can be studied for its effectiveness and ability to improve *is* studied, sometimes in multiple competing ways. Here is just a sample of some of the reviews, appraisals, analyses, and studies that are commonplace in higher education today:

- Entering student surveys like the Cooperative Institutional Research Program's Freshman Survey, administered by the Higher Education Research Institute at UCLA
- Surveys by institutional admissions offices about why students did or did not choose to attend that school
- Course quizzes and tests, including final exams
- Comprehensive examinations for degrees
- Licensure exams
- Self-studies, compliance reports, and inventories for institutional accreditation
- Self-studies, compliance reports, and inventories for specialized accreditation in individual disciplines
- Course evaluations completed by students
- Course evaluations completed by faculty peers or administrators
- Administrator evaluations completed by faculty members, peers, and supervisors

- Staff performance appraisals
- Assessment reports on the effectiveness of the general education program, individual degree programs, and each office or unit on campus
- Comprehensive program reviews to gauge the quality, sustainability, and centrality to mission of various degree programs and campus offices
- Student satisfaction surveys
- Graduating student surveys
- Employee surveys
- Morale studies
- Alumni surveys

Add to these the countless Web sites on which reviews of instructors appear, such as ratemyprofessors.com, rateyourprof.com, ProfessorPerformance.com, and myedu.com. And given enough time, we might come up with several dozen other ways in which professors, academic programs, and institutions are continually reviewed, ranked, surveyed, studied, and assessed. In one sense, it can be misleading to lump very different items together, as I've done in the above list, perhaps leading to the danger of false comparisons. But in another sense, each item represents one important way in which higher education or its critics investigate what people know, how people feel, or what people believe about higher education, all in an effort to determine how well something or someone has performed.

The truly interesting thing is that most of these reviews and studies are relatively new. Colleges and universities seemed to get along without them rather well for centuries, but then, starting in the late 1960s and early 1970s, the use of surveys, inventories, multiple examinations in a course, and personnel reviews began to multiply rapidly. And no matter how different some of us may consider these processes to be, people do tend to confuse them. It's not at all uncommon for faculty members to ask, "Why do we have to do assessment? We already assess our students in class every time we determine their grades," or "Why do we have to

do program review? We already do assessment." In other words, if we really want to get the information we need in order to draw informed conclusions, avoid unnecessary duplication of effort, and improve both our own performance and the effectiveness of our academic programs, we need to understand three things:

1. Exactly what each type of review or evaluation can and can't tell us
2. How to interpret the information gained from that type of review or evaluation
3. Why gathering all that information is important to improvement of our programs and personnel

To obtain that understanding, we have to begin this discussion of faculty reviews and evaluations with a brief history of where all these different processes came from and why they seem to have multiplied so suddenly. The question to ask, therefore, is, "Why in higher education today does everyone always seem to be evaluating everyone else?"

Three trends in American higher education emerged during the 1960s and 1970s that coalesced into what we might term today's account-ability culture in higher education:

1. *The desire of universities to increase their retention rates and levels of student success began causing professors to move away from basing grades largely on a single major project or exam and to introduce multiple grading opportunities throughout their courses.* As the popula-tion of colleges and universities expanded and diversified in the 1950s and 1960s due to the GI Bill and the civil rights movement, many faculty members felt a need to give students earlier feedback about their progress in a course so that they could take action to get back on track, if necessary, before it was too late. Particularly before military conscription effectively ended in the United States in 1973, failing out of college could cause a student to end up being drafted to fight in a highly unpopular war or, at least, to be faced with relatively few options for a desirable career. As a result, higher education made a slow but

perceptible shift from seeing itself as the touchstone that determined who would and who would not graduate to becoming the "student-friendly" or "student-centered" environment familiar at colleges and universities today.

2. *A new theoretical model gave institutions a mechanism for measuring their impact and thus demonstrating to parents, donors, and potential students the benefits that they provided.* Trudy Banta (2002), perhaps the nation's leading expert on assessment and its role in American higher education, credits four major works that emerged between 1969 and 1980 with laying the groundwork for today's culture of academic of academic accountability. First, Kenneth Feldman and Theodore Newcomb's two-volume *The Impact of College on Students* (1969) brought together four decades of research measuring the impact that higher education has on the lives of traditional-aged college students, the type of maturation these students experience during their college years, and the significant role this research could play at the universities of the future. Second, Alexander Astin's *Four Critical Years* (1977) established the metaphor of "value-added" approaches and promoted the use of longitudinal studies to examine net effects. Third, Howard Bowen's *Investment in Learning* (1977) helped establish a public policy context for assessment by emphasizing the societal returns on investment associated with higher education. And, fourth, Robert Pace's *Measuring Outcomes of College* (1979) emphasized the role of college environments and actual student behaviors. Together these works provided higher education with both a conceptual framework for assessment—the goal of college teaching is to improve student learning in mastery of the course's content, critical thinking, and effective communication—and a methodology—the setting and measuring of learning outcomes—that scholars could use to document what a university actually does for students and how effective it is in achieving those goals. Since researchers often act on the principle that if something can be studied, it soon will be studied, it wasn't long before investigations into the measurable impact of higher education began in earnest.

3.   *As the costs of higher education rose despite frequent downturns in the economy, legislators and others who paid the bills for college education began to ask for objective data about their return on investment.* Concurrently with the first two trends, the cost of receiving an education at an American college or university climbed significantly. According to the *Congressional Record* of April 10, 2000, the Senate found that "the cost of attaining a higher education has outpaced both inflation and median family incomes. Specifically, over the past 20 years, the cost of college tuition has quadrupled (growing faster than any consumer item, including health care and nearly twice as fast as inflation) and 8 times as fast as median household income. . . . According to the Department of Education, there is approximately $150,000,000,000 in outstanding student loan debt, and students borrowed more during the 1990s than during the 1960s, 1970s and 1980s combined" (Title I, 2000, p. 5051). Not surprisingly, the number of articles, editorials, and legislative inquiries into the return—frequently in the sense of the economic return—on this investment began to soar as one century ended and the next got under way. Here's how an analysis by the Education Resources Information Center described the situation in 2002:

> The escalating cost of higher education is causing many to question the value of continuing education beyond high school. Many wonder whether the high cost of tuition, the opportunity cost of choosing college over full-time employment, and the accumulation of thousands of dollars of debt is, in the long run, worth the investment. The risk is especially large for low-income families who have a difficult time making ends meet without the additional burden of college tuition and fees. . . . While it is clear that investment in a college degree, especially for those students in the lowest income brackets, is a financial burden, the long-term benefits to individuals as well as to society at large, appear to far outweigh the costs [Porter 2002].

In a similar way, the College Board released a study, *College Pays 2010: The Benefits of Higher Education for Individuals and Society, in Brief* (2010), that documented the difference in median income between

workers with or without a bachelor's degree, the positive effect that universities have on the tax revenues of states and communities, the decreased medical costs incurred by the college educated because of their healthier lifestyles, and so on (Baum, Ma, and Payea, 2010). At the same time, exposés such as Richard Arum and Josipa Roksa's *Academically Adrift* (2011), Andrew Hacker and Claudia Dreifus's *Higher Education?* (2010), Marc Scheer's *No Sucker Left Behind* (2008), and Craig Brandon's *The Five-Year Party* (2010) helped make the public increasingly skeptical that colleges and universities were actually worth their high expense. Internal studies, reviews, and evaluations thus became a way for institutions to document that higher education does indeed make a positive difference in the lives of students and the welfare of the community. Assessment reports and staff evaluations were used to illustrate that colleges were continually focused on improving the quality of their programs, achieving the goals in student learning they claimed to be achieving, and holding faculty members to a very high standard.

Not coincidentally, a lack of public confidence in the quality of American primary and secondary education led to a call for more frequent standardized testing at all levels of instruction. The No Child Left Behind Act of 2001, the Race to the Top Program of 2009 (with its emphasis on uniform standards and assessments), and the frequent demand that teachers' salaries be tied to student achievement have meant that by the time students reach college, they've spent two-thirds of their lives associating education with completing exams, surveys, evaluations, and all other types of assessment instruments as a regular part of the pedagogical process.

Public concern has exacerbated the already growing tendency to test college students more frequently, assess programs more thoroughly, evaluate faculty members more consistently, and review the effectiveness of administrators more rigorously. As a result, it's the rare university today that doesn't have an elaborate set of evaluation procedures and a formal office of institutional research or program

effectiveness and assessment. Faculty members today thus have far greater responsibility for demonstrating to others the benefits of their programs than at any time before the late 1970s. Reviews, evaluations, appraisals, and assessments have become the familiar tools of the accountability culture that pervades American higher education in the twenty-first century. (For more insight into this topic, Larry Braskamp and John Ory, 1994, provide an overview of the factors that brought about regular faculty evaluation in America.)

○ the significance of the accountability culture for reviewers

These background considerations lead us to a major question: While such explanations of why higher education has become so obsessed with evaluation and accountability may have some historical interest, what relevance do they have to a professor or administrator who simply wants to know how best to conduct a faculty review? Or, to put it another way, how does chronicling trends in higher education help anyone become a better evaluator? The answers to these questions may be found in the following principle that will guide us throughout this book:

> You can't review anything or anyone effectively unless you thoroughly understand what you're reviewing, why you're reviewing it, and how the results of your review will be used

In other words, because of the way in which the accountability culture developed in higher education, many different types of appraisals occur simultaneously. These different processes stem from a similar desire but serve very distinct purposes and, as we'll see, the way in which data are collected and analyzed for one of these purposes may make it inappropriate or even impossible to use those data for some other purpose. Why is this so? In order to answer this question, we need to continue our investigation by considering a brief primer on the pro-

cesses that are used to collect information related to the quality of higher education.

The Differences Among Diagnostic, Formative, Summative, and Formative-Summative Processes

One of the key differences among types of review procedures stems from the various purposes for which data are being collected.

Diagnostic processes are designed to gather baseline data about a current situation and, in some instances, to provide insights into the best plan for proceeding in the future. In higher education, a foreign language placement test that determines into which level of study a student should be placed is a familiar diagnostic instrument. SWOT analysis—the identification of an organization's internal strengths and weaknesses, along with its external opportunities and threats—is a diagnostic procedure commonly used in the corporate world, as are inventories that determine the current effectiveness of an existing organizational structure before any modification is implemented.

Formative processes are those that yield constructive advice on how a procedure or performance can be improved. When faculty members survey their students early in a course to determine which pedagogical methods are working well and which need to be improved, they're engaging in a formative process. The students aren't graded on the basis of their answers, and the professor is neither rewarded nor penalized on the basis of what students say; the information merely provides guidance into how to make the class better.

Summative processes are those that result in a conclusive judgment. Assigning students a grade in a course is a summative process, as are promotion and tenure evaluations, academic program reviews, and hiring decisions. At the end of all these processes, a formal decision is made. The student either does or does not pass the course. The faculty member either is or is not promoted. The program is expanded, maintained, reduced, or eliminated. And so on. We might describe the difference by saying that reviewers act as coaches during formative processes; they act as judges during summative processes.

One of the most problematic areas of the faculty review and evaluation system occurs when a process is presented as formative but later assumes a summative role (Manning, 1988). For instance, a class may be told that student ratings of instruction are being administered only to help faculty members improve their pedagogical methods, but results from those instruments are later treated as documentation when the faculty members come up for promotion. The problem with this change isn't merely that the students were misled (although that's bad enough) but also that people tend to respond differently to formative and summative processes. If a class is told, "We'd like to collect some information about how to make this course better for you," the students may come up with a far longer list of complaints and desired improvements than if they're told, "We'd like to collect some information about whether this faculty member will keep his or her job next year" (Stodolsky, 1984; Spencer and Schmelkin, 2002).

When people are aware that an important decision about a person's future will be made on the basis of their answers, they're more likely to keep their criticism within certain limits than when they feel free to speculate about what might constitute an ideal course. For this reason, it's usually important to build a firewall between purely formative and purely summative processes, giving administrators and review committees only information that was collected with the purpose clearly stated. At many institutions, that division hasn't been kept, and this failure can result in appeals, grievances, and even lawsuits after negative reviews. For instance, a 2002 study by the Rutgers University Senate Faculty Affairs and Personnel Committee found that a statement appearing on that school's student course evaluation forms was misleading. With regard to open-ended comments, the forms stated, "This information is intended to be used by the instructor to modify or improve the course," whereas these comments "have been used for personnel decisions in some departments" (senate.rutgers.edu/s0109.html). Blurring the distinction between formative and summative uses of evaluation in this way could create a system in which negative personnel decisions are successfully challenged. When the Rutgers Faculty Senate was

informed about the matter, they promptly removed the misleading statement from the forms.

Nevertheless, we saw a moment ago that "it's *usually* very important to build a firewall between formative and summative processes." There are times when combining these two types of review is either necessary or highly beneficial, although procedures need to be in place to make sure that no one completes an evaluation form under false pretenses. Steven Wininger (2005), professor of psychology at Western Kentucky University, has proposed an approach that he calls *formative-summative assessment* (FSA), which cycles back and forth between the two types of review without misusing the information gathered or misleading those who are involved in the process.

In his own work, Wininger adopted FSA primarily as a pedagogical tool for professors to use with students. For instance, in his educational psychology course, Winniger administers an exam for which students receive individual grades (summative), then uses the exam as a classroom exercise to improve skills where areas of understanding were low (formative), leading to a new grading opportunity (summative), and so on. In a faculty review setting, formative-summative evaluation could consist of conducting a formal evaluation procedure purely as a developmental exercise, with the stipulation that the results may not be considered during any later reviews, using the results of that procedure to develop a strategy to remediate any perceived areas of weakness, and then conducting the formal procedure again, this time in a summative manner, when a personnel decision (such as contract renewal, merit increase, or promotion) needs to be made.

In fact, in many systems, probationary pretenure reviews, which we'll consider in Chapter Five, are intended to function in precisely this way. The pretenure review provides the formative component of the process, while the actual tenure evaluation two or three years later provides the summative element. The complicating factor is, as we shall see, that faculty contracts are sometimes not renewed after particularly poor pretenure reviews. It's important, therefore, that the probationary faculty member who's undergoing pretenure review understand precisely what

the possible results of the process may be. Otherwise even the best-intended formative-summative process could create problems for both the reviewer and the institution as an entity that failed to follow its own policies.

The Differences Among Review, Appraisal, Assessment, and Evaluation

The distinctions among diagnostic, formative, summative, and formative-summative reviews are important because they help to explain why higher education has developed so many processes that appear to gather very similar information. But there's also another reason for the multiplicity of these processes: due to the accountability culture in which higher education operates today, stakeholders want to verify that all parties involved in the academic enterprise are meeting the standards that have been set for them. The faculty and administration want to make sure that students have reached a certain level of performance before issuing a diploma. Accreditation bodies want to make sure that programs are adhering to certain requirements and guidelines before certifying them. Legislatures and governing boards want to make sure that institutions are using resources effectively before authorizing a new budget. Students and parents want to make sure that they're getting their money's worth. Because different types of conclusions need to be drawn by these different stakeholders, it's inadvisable—and frequently impossible—to make a single process that achieves all these purposes simultaneously. As a result, it's important for anyone involved in these processes to keep the terminology straight so as to avoid distorting the results or confusing the participants.

Although the terms are used with different meanings by other authors or in other situations, it's useful to clarify what these words are intended to mean in this book. For our purposes, a *review* is the most general term for any study that examines the performance of an individual or organization. We can thus refer to all processes we'll be considering—appraisals, evaluations, and assessments—generically as reviews. For this reason, every evaluation is a review, but not every review is an evaluation, and so on.

An *appraisal* is a type of review in which someone in a position of authority determines whether a person, unit, or program is meeting an established set of expectations. For example, in a performance appraisal, a supervisor rates an employee according to the degree to which he or she is handling the responsibilities detailed in that person's job description. In a program appraisal, the director or chair rates the degree to which a program is fulfilling its key objectives. In this way, appraisals always exist within a hierarchical reporting structure and involve comparisons between anticipated standards and actual performance. They're a much more focused type of review than some of the others we'll explore. In academic settings, it is most often administrators, staff members, and relatively small programs that undergo appraisals instead of other types of evaluation. Faculty members, complex programs, and the institution as a whole may undergo appraisals from time to time, but they are more frequently reviewed in other ways.

It is the difference between *assessment* and *evaluation* that tends to cause the greatest confusion on most college campuses. Perhaps the best way to understand the difference that I'll adopt in the chapters that follow is to consider how these processes diverge in three key areas (Buller, 2010, p. 88):

ASSESSMENT IS	EVALUATION IS
Formative; provides constructive advice on how an activity can be improved	Summative; renders a judgment as to whether a specific action is warranted
Process oriented; looks for ways to improve the activity itself	Goal oriented; determines whether a particular level of quality has been reached
Primarily concerned with larger or collective entities, such as an entire group of students in a course, courses in a curriculum, or departments in a college	Primarily concerned with individual entities, such as a single person, course, or program

Other books and systems use these terms in different ways—and you'll undoubtedly have your own preference—but for our purposes, I'll adhere to the distinctions I have outlined. They offer a convenient way of making some significant distinctions among various kinds of reviews. The key to these differences may best be seen by returning to the common faculty questions we encountered earlier: "Why do we have to do assessment? We already assess our students in class when we determine their grades." In light of the differences just outlined, grading a student in a course is an *evaluation*: it tells an observer how well that individual student did in meeting the course's pedagogical goals. In this manner, assigning a grade is summative: it renders a judgment about the level at which that student performed.

But student grades reveal nothing at all about the effectiveness of the course itself or how well that course has been integrated into an overall program of study. Those grades are influenced by too many other factors: each student's individual ability, how hard he or she studied, his or her familiarity with the material from previous courses, and so on. Moreover, nearly everyone in higher education is familiar with a situation in which the students of an introductory course all achieved very high grades but then performed poorly in later courses or on licensure exams. In this case, the students succeeded in the individual evaluations they'd received in their introductory courses—their tests, quizzes, papers, and assignments—but the curriculum as a whole hadn't been very effective in preparing those students for their later needs.

In order to gain a sense of how whole programs can be improved, colleges and university use *assessment*, an activity designed to improve the process of learning found in that program, not to rate the performance of individual students. The word *assessment* is derived from the Latin verb *assidere*, meaning to "sit beside." We can imagine assessment as what trusted counselors or mentors do when they sit with us, giving us the benefit of their guidance. *Evaluation* comes to English through the early French *evaluer*, which combined the Latin words *ex* (out of, from) and *valere* (to be strong, to have strengths) to denote the act of extracting something's merit from it or assigning it a value. Schools

sometimes combine these two processes by embedding certain questions or types of problems into the exams of a particular course, but in this case, the intention of the two activities is different. The entire exam serves to evaluate each student; the same question on all the students' exams serves to assess the course. But no student's grade or even the average of all grades really tells you very much about how well a course has been designed or how effectively a professor has been teaching. Individual students may perform well, poorly, or somewhere in between in even the best-designed program, and any single student may excel in even the least effective program. You can't draw evaluation conclusions from assessment or vice versa.

○ why these differences matter when reviews are conducted

When we set out to conduct an actual review, the differences we've examined in the last two sections become very important. For example, if we don't distinguish carefully among evaluation, assessment, and appraisal, we can end up recommending unnecessary revisions to the curriculum when the problem is better addressed through faculty development or student remediation. Similarly, if we are careless in combining formative and summative processes, we could have our decisions overturned or find ourselves being sanctioned for failing to follow appropriate procedures. For this reason, we should begin every review process by making sure that we clearly understand the answers to the following questions:

1. What is the specific purpose of this review?
2. Who or what is being reviewed?
3. What are the possible outcomes of the review? In particular, what are the possible outcomes if the result of this review is strongly negative?
4. Is the person, unit, or program undergoing the review aware of these possible outcomes?

5. With whom will the results of the review be shared?
6. What sources of information am I required by institutional policy to consult as I conduct this review?
7. What sources of information am I allowed—even though not required—by institutional policy to consult as I conduct this review?
8. Are there strict deadlines I must meet in order for the results of my review to become valid?
9. What is the nature of my conclusion? Am I authorized to make a decision, or am I merely recommending a course of action to a supervisor, committee, or other party?
10. Does institutional policy require me to retain records of how I reached my decision for a certain period of time?

By understanding these issues before a review gets under way, we're less likely to collect or consider the type of information that doesn't really help us, make the type of mistakes that will invalidate our efforts, or have our own findings overturned on appeal or by higher administrative levels. In later chapters, we'll also consider other issues that should be considered for different types of reviews. But the central principle is always the same: *Unless you know precisely why you're conducting a review, you will have a very hard time identifying what to review and how to review it.*

○ why review and goal setting should be merged into a single, seamless process

Review by its very nature is a retrospective process. Even tenure reviews, which are intended to identify a faculty member's promise for the future, are based on a consideration of past achievements. But it's important that other types of review not become so focused on the past that they lose all sight of the future. Nearly every review should end with an agreement about specific goals for future development. It's easy to understand why goal setting would be important in formative evalua-

tions. After all, the whole purpose of the process is to advise the person or program on ways to improve.

But why would goal setting be important for a summative process? Since these reviews culminate in a decision, what is the point of looking toward the future? The answer to these questions is that even summative processes don't imply that no further growth or achievement is possible. A faculty member who is promoted to the rank of associate professor still needs to earn promotion to the rank of full professor, and a faculty member who is promoted to the rank of full professor still needs to be successful in future posttenure reviews. A curriculum that receives universal praise during a program review will still be reviewed again in the future, and course content is always in a state of evolution and improvement as students change and knowledge expands. For this reason, review processes should conclude whenever possible with a discussion of what's next and the setting of future goals. These goals shouldn't restrict the person or program from taking advantage of unexpected new opportunities or modifying his or her plans as circumstances change. Nevertheless, the goals provide a general road map for further progress and a starting point for measuring that progress during future reviews.

You will notice that the language I have used is, "Review processes should conclude *whenever possible* with a discussion of what's next." When would it ever not be possible to merge reviews and goal setting into a single, seamless process? In certain systems, when a tenure decision is negative, the institution is forbidden to provide reasons for that decision. The rationale is that during a faculty member's probationary period, the institution is allowed to forgo renewing his or her annual contract for any reason or for no reason. In fact, the necessity for an institution to provide a reason for nonrenewal of contract is (along with the right of due process to challenge that decision) the very essence of what tenure guarantees. As a result, some systems consider it inappropriate to extend that right to an untenured faculty member who has failed a review by offering an explanation for the reasons behind the decision. (See, for example, Lander University, 2010; University of

Louisiana at Monroe, 2007; University of Tennessee, 1981.) Setting goals and developing plans for improvement with the faculty member could be construed as revealing why tenure was denied; looking toward the future thus unnecessarily clouds the true meaning of a negative tenure decision and opens the institution to a possible legal challenge. Since technically no justification is required, certain institutions conclude, no justification should ever be provided.

conclusion

A valuable source for emerging approaches to faculty review is *Assessment and Evaluation in Higher Education*, a peer-reviewed journal published seven times a year by Routledge. Known as *Assessment in Higher Education* until 1981, the journal's current title reflects its focus on both major types of review. Each issue contains eight to ten articles dealing with such issues as best practices in the use of student ratings, techniques for evaluating adjunct or distance learning faculty members, assessment techniques for measuring student progress in mastering new technologies, and so on. Since issues in higher education develop so rapidly, it's a good idea to supplement insights gained through books with an understanding of emerging trends, reported in such publications as the *Chronicle of Higher Education* and *Assessment and Evaluation in Higher Education*.

References

Amano, I. (1990). *Education and examination in modern Japan*. Tokyo: University of Tokyo Press.

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago: University of Chicago Press.

Astin, A. W. (1977). *Four critical years*. San Francisco: Jossey-Bass.

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

Baum, S., Ma, J., & Payea, K. (2010). *Education pays 2010: The benefits of higher education for individuals and society*. New York: College Board. Retrieved from trends. collegeboard.org/downloads/Education_Pays_2010.pdf.

Bowen, H. R. (1977). *Investment in learning: The individual and social value of American higher education*. San Francisco: Jossey-Bass.

Brandon, C. (2010). *The five-year party: How colleges have given up on educating your child and what you can do about it*. Dallas: BenBella.

Braskamp, L. A., & Ory, J. C. (1994). *Assessing faculty work: Enhancing individual and institutional performance*. San Francisco: Jossey-Bass.

Buller, J. L. (2010). *The essential college professor: A practical guide to an academic career*. San Francisco: Jossey-Bass.

College Board. (2010). *College pays 2010: The benefits of higher education for individuals and society, in brief*. New York: College Board. Retrieved from http://trends. collegeboard.org/downloads/Education_Pays_2010_In_Brief.pdf.

Feldman, K. A., & Newcomb, T. M. (1969). *The impact of college on students*. San Francisco: Jossey-Bass.

Hacker, A., & Dreifus, C. (2010). *Higher education? How colleges are wasting our money and failing our kids—and what we can do about it*. New York: Times Books.

Kehm, B. (2001, March). Oral examinations at German universities. *Assessment in Education: Principles, Policy and Practice*, *8*(1), 25–31.

Lander University. (2010). *Faculty handbook*. Retrieved from www.lander.edu/Libraries/ Site_Documents/Faculty_Handbook.sflb.ashx

Manning, R. C. (1988). *The teacher evaluation handbook: Step-by-step techniques and forms for improving instruction*. Upper Saddle River, NJ: Prentice Hall.

Min, H., & Xiuwen, Y. (2001, March). Educational assessment in China: Lessons from history and future prospects. *Assessment in Education: Principles, Policy and Practice*, *8*(1), 5–10.

Pace, C. R. (1979). *Measuring outcomes of college: Fifty years of findings and recommendations for the future*. San Francisco: Jossey-Bass.

Porter, K. (2002). The value of a college degree. *ERIC Digest*. Washington, DC: ERIC Clearinghouse on Higher Education. (ERIC Document Reproduction Service No. ED470038). Retrieved from www.ericdigests.org/2003–3/value.htm

Scheer, M. (2008). *No sucker left behind: Avoiding the great college rip-off*. Monroe, ME: Common Courage Press.

Smallwood, M. L. (1935). *An historical study of examinations and grading systems in early American universities*. Cambridge, MA: Harvard University Press.

Spencer, K., & Schmelkin, L. P. (2002). Student perspectives on teaching and its evaluation. *Assessment and Evaluation in Higher Education*, *27*, 397–409.

Stodolsky, S. S. (1984). Teacher evaluation: The limits of looking. *Educational Researcher*, *13*(9), 11–18.

Title I: Levels and Amounts, Sec. 302. (2000, April 10). *Congressional Record*, 146, Pt. 4.

University of Louisiana at Monroe. (2007). *Faculty handbook*. Retrieved from
    www.ulm.edu/ncate/standard%205/Exhibits/5d1.1/FacultyHandbook.pdf.

University of Tennessee. (1981, May 25). *Faculty senate memorandum*. Retrieved from
    web.utk.edu/~senate/MemoReasonsDenial.html.

Wininger, S. R. (2005). Using your tests to teach: Formative summative assessment.
    Teaching of Psychology, *32*, 164–166.

Wiseman, S. (1961). *Examinations and English education*. Manchester: Manchester Uni-
    versity Press.

Resources

Buller, J. L. (2008). Why all evaluations aren't alike. *Academic Leader*, *24*(12), 4–5.

Feldman, K. A. (2007). Identifying exemplary teachers and teaching: Evidence from
    student ratings. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and
    learning in higher education: An evidence-based perspective*. Dordrecht: Springer.

Goldman, G. K., & Zakel, L. E. (2009, May). Clarification of assessment and evaluation.
    Assessment Update, *21*(3), 8–9.

Langen, J. M. (2011). Evaluation of adjunct faculty in higher education institutions.
    Assessment and Evaluation in Higher Education, *36*, 185–196.

Theall, M. (2010). Evaluating teaching: From reliability to accountability. *New Directions
    for Teaching and Learning*, *123*, 85–95.