

Evaluando la calidad de la Universidad – Particularmente su enseñanza

Assessing the Quality of a University-Particularly its Teaching

Robert E. Stake

University of Illinois at Urbana – Champaign
(United States)

María Isabel Arbesú García

Universidad Autónoma Metropolitana
Unidad Xochimilco (México)

Gloria Contreras Pérez

Pontificia Universidad Católica de Valparaíso
(Chile)

Robert E. Stake

University of Illinois at Urbana – Champaign
(United States)

María Isabel Arbesú García

Universidad Autónoma Metropolitana
Unidad Xochimilco (México)

Gloria Contreras Pérez

Pontificia Universidad Católica de Valparaíso
(Chile)

Resumen

Las universidades se constituyen en organizaciones complejas compuestas por estructuras semiautónomas y por un profesorado aun más autónomo. La tradición y las normas dirigen las actividades hacia la homogeneidad. Así y todo, la mayoría de administradores, profesores y alumnos gozan de un gran margen de libertad para desarrollar su trabajo. Vincent

Abstract

Universities are complex organizations, made up of semi-autonomous sectors and even more autonomous faculty members. Tradition and rules push activities toward homogeneity. Still, most administrators, staffs and students enjoy a grand leeway in which to work. Vincent Tinto (1995) noted that it would be difficult to find a single academic culture but, different

Tinto (1995) señala que es difícil identificar una cultura académica única, ya que existen diferentes culturas, diferentes prácticas de evaluación y diversidad de grupos en cuanto a profesores y alumnos. Abundan los procedimientos formales e informales de evaluación y todos ellos, individual y colectivamente, resultan insuficientes a la hora de generar una valoración completa de la calidad universitaria. La mejor comprensión de la calidad podemos encontrarla en aquellas personas con experiencia que atienden especialmente a los flujos de información que se dan en el campus y que rechazan los intentos por hacer de calificaciones y escalas superficiales elementos representativos de la calidad del conjunto o de alguna de sus partes. Quien más interés tiene en la evaluación, aunque pocas veces afronta con franqueza la calidad educativa, es la administración, que por supuesto necesita evaluaciones internas y externas. La evaluación de la enseñanza en el campus es un ejemplo especialmente cercano de la complejidad y de los métodos de evaluación, y constituye la referencia principal en este documento. La distinción entre evaluación formativa y sumativa es crítica, y la misma información rara vez sirve para ambas. Se dice que las estrategias del uso formativo de datos para mejorar la universidad son: la estrategia basada en objetivos, la basada en el déficit, la basada en un sector de estudio y la estrategia basada en el autoestudio, estrategias que se solapan y complementan, consiguiendo los mayores beneficios cuando se consigue poner en juego el potencial de la estrategia del autoestudio.

Palabras clave: evaluación; calidad de la enseñanza; calificación de universidades, evaluación formativa.

cultures, different practices of evaluation and various groups of teachers and students. Informal and formal evaluation procedures abound, and all, individually and collectively, fall far short of delivering a comprehensive statement of university quality. The greatest understanding of quality is to be found in experienced individuals who attend to the flows of information across campus and disdain efforts to make shallow ratings and reputations stand for the goodness of the whole or certain parts. The most invested in evaluation yet seldom forthright about educational quality is the administration, which of course needs both internal and external evaluation. The evaluation of teaching on campus is a particularly accessible example of complexity and assessment methods, and serves as the primary example of this document. The distinction between formative and summative evaluation is critical, and the same data seldom can serve both. The strategies for formative use of data to improve the university are said to be: The goal-based strategy, the deficit strategy, the sector-study strategy, and the self-study strategy, which overlap and complement, but get the greatest gain when the energy of self-study can be brought forth.

Key words: assessment, quality of teaching, university ratings, formative evaluation.

Diverse assessments across the university

The buildings wear away, but mostly they fail the new burdens put upon them, such as for computing and collaborative instruction. The university's research seeks breakthroughs, with funding for inquiry problematically distributed (Alpert, 1995). For students, the university is partly a country club, not only the pools but with cultural exclusivity. The promises and expectations of the university vary greatly from corner to corner, and each begs for sensitivity and repair.

Assessing the quality of campus work, programs, and facilities is a responsibility shared by a vast array of stakeholders, including the trustees, the central administrators, the sector administrators, and of course, the faculty and the unions, and the students and their families, the legislature and other funding bodies, the workplace, the media and more. They all evaluate what the university does, mostly informally. Much of what the university does is complex and invisible. The great bulk of assessment is informal, structured by introspection and personal experience and framed against the long-standing reputation and culture of the university (Dressel, 1971).

The total quality of any university is irregular, conditional and as complex as its many functions, classrooms, libraries and laboratories. To rate a university as seventeenth or in the top ten or even as "very good" is a gross oversimplification, no matter the specificity of criteria or the length of acquaintance. Rankings are commonly voiced because people are comfortable with stereotypes and the reflected glow of self-adulation.

Administration. Assessment is important for managing the university and its parts. Call it assessment, call it evaluation, call it quality control, it can improve understanding when considered against thoughtful expectation and realistic standards.

Comparison of universities or departments serves curiosity but serves little the managerial or public good. The more valuable assessments are those that are formative, developmental, aimed at avoidance of dysfunction, repairing weakness, and shaping new teaching, research and public service.

As expressed by its administrators, individual universities have their character—a mission, an organizational health, an ethic. Each has its uniqueness, integrity and diversity. Each department, professional program, and area of aesthetics contributes to that character. Each individual in the university community, knowingly and unknowingly, contributes. The aggregated strengths and weaknesses of the whole and of the parts are difficult to know. Paying close attention to: events, the flow of operational data, and the mood of the community, and more, yields good indicators. But guiding the largest and smallest of university functions depends less on formal indicators and more on experience and introspection.

There is much room for improving assessment of quality at universities nearby and around the world. There is considerable danger in representing the complexity of operations by simple indicators and criteria. Against this background, there are many topics to consider but we will limit this paper to four, one of which will be to the assessment of the special role of teaching.

Purposes. The purposes of this paper are (1) to indicate the tension between summative and formative assessment at the university, (2) to identify barriers to comprehensive assessment at the university, (3) to overview the responsibility for assessment of university teaching as an example of accountability for the merit and the shortcoming of university functions, and (4) to analyze the role of formal and informal assessment in improving such operations as teaching.

Formative and summative assessment, formal and informal

People everywhere take notice of what they do, seeing in it the good and the bad. Evaluating is part of everyone's normal behavior. On the university campus, almost everything is evaluated, by the people in charge, by the direct beneficiaries, sometimes by the public (Stake & Cisneros, 2000). The great portion of evaluation is informal, even unintentional, whether done by administrators or others. On some occasions, and as part of some operational routines, extra effort is made to observe, measure, record, and interpret the merit and shortcoming of things. When evaluation becomes planned, routinized, recorded, and publicized, we say that the assessment is formal. Many of the functions of the university are assessed both formally and informally.

A rationale for formal evaluation should determine methods, procedures, instruments and activities. Often there are several, sometimes competing, rationales for evaluation of any university entity or function. (Whatever is being evaluated, some call it an "evaluand.") Especially important is the anticipated use of the evaluation. People are not going to act the same with formative and summative evaluations (Contreras, 2010).

Summative evaluation. When evaluation is done primarily to understand the quality of an existing evaluand, we call it summative evaluation. We think of it as having value, knowing that the value depends on at least somewhat its situation, and its potential uses and the users themselves, but for the time being, concentrating on how good or how bad it is. When we speak of a university's collaboration with industry, or its music program, or its library, we usually are speaking summatively of its quality at a particular time and place.

Formative evaluation. But often we see the evaluand more as something being developed, changed, reorganized for new use--and we call for formative evaluation. The arborist trims the tree, the counselor changes protocol, the coach recruits an athlete, all of them evaluating formatively how to improve the process. In formative evaluation, not only the prospective use is considered, but also the likely situations, the new opportunities, the barriers and possible disasters.

It would be wrong to think that summatively assessing university quality gives no attention to change and utility. And wrong to think that formative assessment gives no attention to the evaluand's merit in its immediate form. Some mix of formative and summative evaluation is common. But the different purposes call for different criteria, different explanations to informants, and different instruments.

University and department accreditation by professional associations is primarily summative assessment, but philosophically, and occasionally operationally, it is a process for improving the institution (Dressel, 1971). Some countries rely heavily on accreditation,

others on inspection. Some countries rely on standardized quality control procedures set up by the central governments. Failure to attain accreditation is sometimes an indication of non-conventional operating procedures more than serious weakness. Even with thoughtful rationales, the validity of what is actually accredited is something to review. Any summary score (such as “accredited”) oversimplifies the strengths and weaknesses of the evaluand.

A hypothetical conversation between administrators

A1: We are thinking of dropping accreditation by two of the three associations now accrediting us.

A2: Won't it hurt your reputation?

A1: I don't know. We agree with most of their standards, and being accredited may lower the howling, but it's too expensive.

A2: Don't you learn from their assessments?

A1: Mostly they reinforce what we already know. Sometimes we learn more from our own committees. Two of them last year really studied the problem.

A2: Committees can be worthless. And they sap staff energies.

A1: And they do little to help our reputation in the media.

A2: I suggest you find better ways to use the accreditation review process. Maybe your people show little respect for accreditation, so they treat it as political rather than evaluative.

A1: I don't know. We have to cut expenses.

Barriers to comprehensive assessment across the university

In previous paragraphs, the university is said to be complex and far from homogeneous. Its parts and people have autonomy with different histories, aims, and operations. Nearly impossible, but still we do evaluate--and should comprehensively. Comprehensive assessment is illustrated in the following section, an example on assessment of teaching. It does not say to arrive at one summary rating but to consider the multiple parts, their multiple features and functions, and multiple ways of knowing them. Common barriers to assessment are to be found in every institution. The order of barriers below does not indicate an order of importance.

1. Widespread opposition to formalized assessment. Although university people regularly evaluate their activities informally, for example, in conversations with administrators and students, often they do not trust formal evaluation to treat them fairly. This attitude has many origins: popular distrust of authority, concern about oversimplification, lack of broad experience, doubts about the use of the information, and others.

2. Misunderstanding of administrators of formal evaluation. Administrators frequently associate formal evaluation with structured strategy, measurement, objective data gathering, statistical analysis and indicators--all for control of the activities at the

university. Often they have to show results quickly, so sometimes they favor techniques that capture a great amount of information in a short time.

3. Lack of funding and training for people evaluating the functions. Although evaluation is increasingly supported by many institutional centers at least partly dedicated to evaluation and improvement of quality, frequently staff members have only limited knowledge about the evaluation process, for example, to design studies, to fine-tune instruments, to collect data, to analyse information or to report results. Universities do not have funds for comprehensive formal assessment of the functions.

4. Political implications. Evaluation is seldom recognized for its political nature, finding constraints on what will be studied, reward for favorable findings, and protective control of distribution of information. (Cronbach, 1977)

5. Expectation that the same data can be used for quite different purposes. Data are collected from people expecting single uses, such as for research, but it ends up being used more widely. The meaning of the information is often not the same for summative and formative uses. Seeing this happen causes misinterpretation and loss of trust.

6. The reactive effects of assessment. Even without intention, the process of evaluation results in changing behavior. When a criterion becomes known as important for decision-making, people consciously and unconsciously, sometimes unethically, act in such a way as to get better scores and improve their situations.

7. Ethical ownership of personal data. The confidentiality of information, especially for the people being evaluated, is a complex issue. In many ways, the institution is respectful of the privacy of personal information but, in others, the authorities use private information to support decisions against those persons.

8. Self-Promotion. The increasing obligation for individuals, sectors, and the university as a whole to boast of strengths and accomplishments makes it important not to admit weakness and lack of accomplishment. Thus, assessment is denied important information needed.

Some barriers are especially apparent in the matter of evaluation of teachers, (which is the topic of the next section of this paper). The barriers above are pertinent to general assessment, including that of teaching. The barriers below, specific to teaching, are clues to obstacles to quality control throughout the rest of the university:

9a. Conceptions of teaching and evaluation of teaching. Every instructor at the university has his or her own ways of teaching, and these ways are related to the discipline, previous professional and personal experiences, particular contexts, particular students, and some other aspects. Thus, a university may have many different conceptions about teaching and evaluation of teaching, but there is common meaning throughout, as well.

9b. Undervaluation of teaching. Although it is said that teaching at the university is very important, it has been regularly undervalued compared to many other functions

like research or public service (Canales, 2003). This situation has led to a lack of attention and resourcing of teaching, with a lack of guidelines for its evaluation.

9c. Invalid indicators. Indicators of teaching quality may appear to have validity, such as “clarity of speaking,” and may, for groups, be positively correlated with teaching effectiveness, but for individual instructors those criteria are not valid for administrative personnel decisions.

These several barriers can be overcome but they are strong and pervasive. They manifest differently in different situations. Continued study of them should be part of the institutional research of the university. Educational researchers are well positioned to study these barriers for the assessment of teaching, partly as examples of how evaluation is frustrated in all sectors of the university.

An example: The assessment of teaching

Evaluation of teaching is one of the most important assessments at the university. In this section, the exclusive focus will be on assessment of teaching. Each assessment of other university functions has its special character but the points made in this section have relevance for evaluation of the other parts of the university. Assessment is a search for work needing protection from undue reorganization, reduction, and modernization. And assessment is a search for flaws needing fixing (Lehman, 1975).

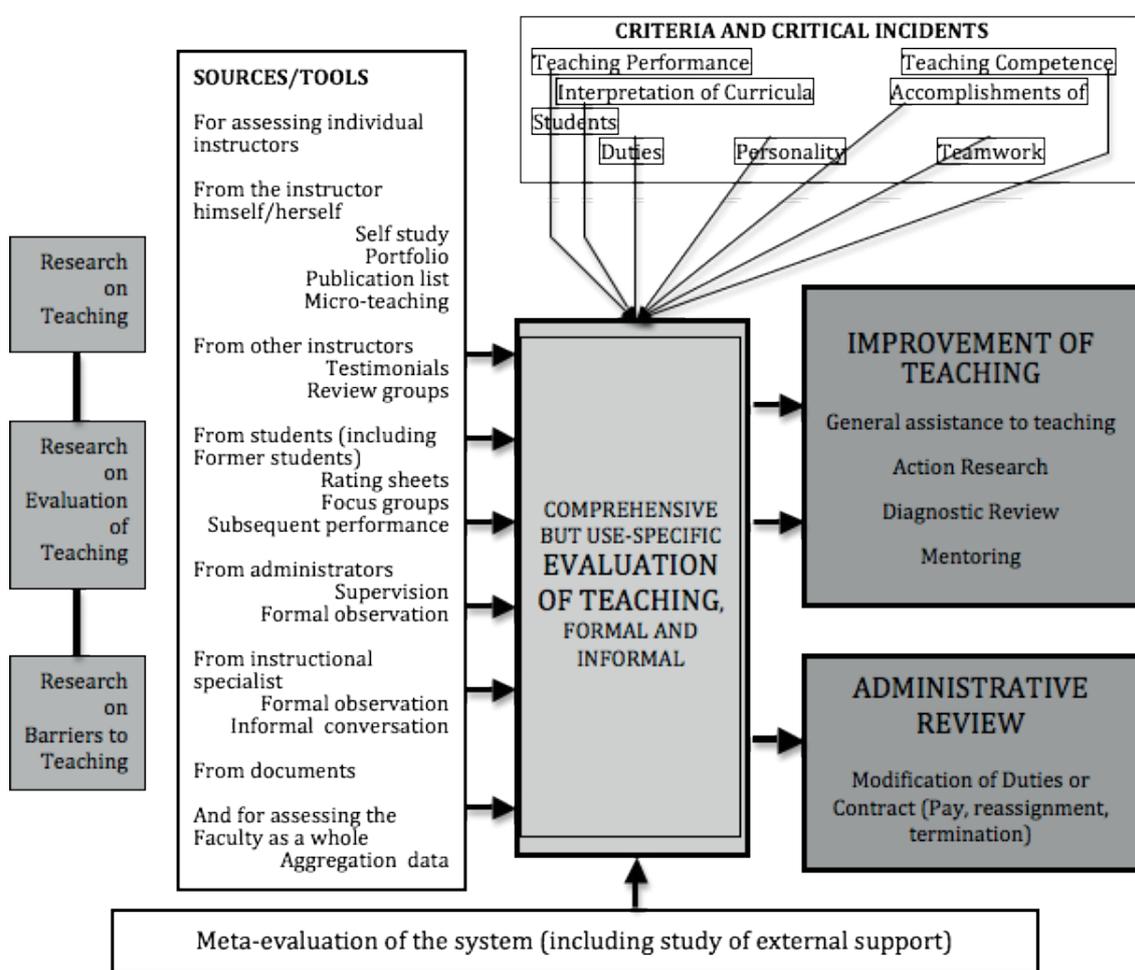
Given current formal assessment procedures, a university’s teaching, as a whole, or that of a department, can almost only be assessed by some aggregation of individual instructor teaching quality. Holistic self-study and committee review are possible, but direct assessment is outside reality. Good assessment of teaching quality of individuals is possible, and that is the concentration of this section.

The complexity of teaching. Teaching is a complex human activity fitted to common but ever-different situations. It is expected to result in student learning, but that has multiple definitions. Harry Broudy (1963) identified the modes of teaching as didactic, heuristic, and philetic. Philetic is establishing a relationship between teacher and student, resulting in motivation and perhaps love of learning what the teacher has to teach. The didactic is direct teaching, the imparting of knowledge and skill. Indirect teaching is the arrangement of circumstances so that students become better educated. Good teaching includes interpreting the curriculum, inspiring, scheduling, making assignments, supervising classroom interaction, testing, grading, advising, collaborating with peers, corresponding with administrators, and more (Darling-Hammond and Young, 2002; Arbesú, 2006; Arbesú, Díaz Barriga, Elizalde, Luna, Rigo, Rueda & Torquemada, 2007; and David, 2010). Doing one thing well does not assure doing all things well. Situations will vary as to how different responsibilities will be weighted.

Some teaching is open to view but most university teaching is private within the classroom. By tradition, many instructors have come to expect privacy in the classroom and would object to having unidentified observers or electronic monitors in the room. With instructors feeling secure and protected well beyond “academic freedom,” their privacy probably contributes to effective teaching. Sometimes it hides incompetent teaching. Administrators and other stakeholders of the university have a right to know

the content and methods of each instructor, but monitoring greatly diminishing the privacy of teaching would probably be an obstacle to the improvement of teaching.

Communitarian teaching. The traditional concept of evaluating teaching is the evaluation of an autonomous instructor in an individual classroom. An additional perspective is found in practice, assessing the contribution an instructor makes to the maintenance and improvement of the instructional programs of the department. What instructors contribute to the integrity of all offerings, not just their own, is important too. A charismatic lecturer or innovative lab-organizer or personalistic mentor, that is, a star, sometimes contributes little to the upgrade of weak, misdirected, frivolous and out-dated courses in the department. Both individual and team contributions need to be considered when teaching is assessed.



Source: Prepared by the authors.

Figure 1. Representation of the Comprehensive Evaluation of Campus Teaching.

Collaboration across a campus faculty about matters of teaching is not new (Cruz Souza, 1999), but, in most places, it remains the exception more than the practice. Writing about a faculty as a “community of practice” became identified with John Seely Brown (1997) and others at the Institute for Research on Learning in Palo Alto (Community Intelligence Labs, 1997). Colleague Etienne Wenger said: “Even those who speak about learning organizations, lifelong learning, or the information society do so mostly in terms

of individual learners and information processes. The notion of communities of practice helps us break this mold (1991:7).”

A representation of the evaluation of teaching

Figure 1 is a graphic representation of both the formal and informal evaluation of teaching in higher education. It is more a representation of responsibility than a guide to assessment. It displays the responsibility for assessment of university teaching as an example of accountability for merit and shortcoming of all major university functions.

The shaded box in the middle represents responsibility for the evaluation of teaching, and to the right, responsibility for the improvement of teaching. Although potentially integrated, here they are kept separate because they are different responsibilities. At times they are merged as a comprehensive formative evaluation responsibility. For “Improvement of Teaching,” we show four main areas of university support, (1) to provide counsel on improvement, (2) to foster personal self-study, (3) to provide measurement services regarding teaching effectiveness, and (4) to support mentoring by someone more experienced in teaching within the discipline. There are more, of course.

The box next below indicates the separate track to administrative review for the purpose of quality control--by changing instructional practice broadly as well as by modifying an individual teacher’s contract. Some evaluation, summative evaluation, serves administrative review of teaching, as indicated. Much of the data are “use-specific,” intentionally gathered for a specific purpose,. All too often, any one measure will not serve both formative and summative purposes. In fact, data intended to improve teaching should not be automatically available to administrators because assistance to instructors can be more effective when information on weaknesses is kept private.

The range of data for evaluation is shown at the left. It comes from several sources, usually from the teacher being evaluated and his or her administrators, but also from peers in the department, from students, and from specialists in instruction. It starts with teacher reflective self-study but ranges much further (Brubacher, Case, and Reagan, 2005). Using only one or two of these sources is likely to yield a narrow view of the quality of teaching. At the top right are the criteria and critical incident data to be described in the next section. These responsibilities for evaluation need to be carried out against a background of research and professional experience, including what is pertinent to each particular university.

Some evaluation of teaching (at the bottom of the Sources/Tools box) pertains collectively to the department or sector, whether by aggregation of teaching performance of its instructors or by studying it as a unit. The three boxes at the left indicate the responsibility of drawing from research on teaching and on personnel evaluation, as well as research on barriers to teaching (mentioned earlier). The lowest box indicates the enduring necessity of backing up all evaluation with meta-evaluation that is, evaluating the evaluation (Stufflebeam, 1981).

To repeat, this chapter is not a guide to evaluation practice but the identification of responsibilities shared by individuals, groups, and the university as a whole. Many

of these responsibilities will be engaged in informal evaluation. Inadequate resources for formal evaluation and assistance for improvement of teaching will continue to limit engagement in these numerous obligations, but recognition of the error of using one or two indicators will be important. The criteria and gathering of critical incidents should be comprehensive.

Criteria and critical incidents. Two grounds for evaluating teaching come from quantitative and qualitative perspectives: criteria and critical incidents. In a sophisticated system, both will be given attention. A criterion is a distinguishing property or characteristic of any thing by which its quality can be judged or estimated, or by which a decision or classification may be made (Sadler, 2005). A criterion is often represented by a scale running from none to much. Good formal evaluation is often organized around multiple criteria drawing upon multiple data sources.

A critical incident is a distinguishing event or condition by which quality can be judged or estimated (not to be confused with the critical incident technique, Flanagan, 1954). Sometimes a single critical behavior or utterance is so important that it may outweigh other virtues and may be the primary basis for action, such as dismissal (or for an institution, loss of accreditation). In 2008, in two separate incidents, the University of Iowa removed a professor from his classrooms and duties after he had been charged by a student with misconduct. Isolated for a long period, each committed suicide (Franke, 2008). Each incident was a consideration for evaluating the university. Of course other incidents will be positive, usually ordinary, such as a teacher stepping forward to defend someone. In a report, such incidents are represented by experiential detail in context rather than by measurement.

The appropriateness of criteria and critical incidents for evaluating depends on the purpose of the evaluation, the people who are evaluating, the institution, the discipline being taught, and so on (Cisneros-Cohernour, 1997; Canales, 2003). Criteria and critical incidents can be used together but will seldom converge on single pictures of teaching quality.

The most common criteria and types of critical incidents for evaluating university teaching are:

- Teaching performance. To judge performance is to judge what the instructor actually did.
- Teaching competence. To assess competence is to measure or estimate what the instructor is capable of doing, for example, by interview and review of portfolio, to assess the teacher's understanding of the subject matter being taught.
- Curriculum interpretation. To judge teaching as to the way that teachers prioritize and interpret the topics contained in the curriculum. Here we have for example, the capability to complete all lessons of the syllabus and to be constrained in offering objectionable interpretation of matters.
- Duties. Criteria are fixed a priori by contract, assignment or tradition on what the teacher is obligated to do. Here we have aspects like being present at classes and preparing lesson plans (Scriven, 1995).

- **Personality.** The criteria are based on certain ideal characteristics of a teacher, like empathy, sense of humor or relationship with the students,
- **Teamwork.** Teamwork is a special competence, personality trait and commitment to work in a collaborative way with others, particularly other members of the faculty.
- **Students' achievement.** One way to judge teaching is based on student grades, projects, performance, or scores on standardized tests. One might look at value added as seen in comparison of student test scores across years (Rosu & Stake, 2008; Rothstein, 2008). The aptitude of the students is an important consideration.

Whatever criteria or critical incidents are used, people responsible for evaluation of teaching seek widespread understanding of the means to judge teaching. This remains a problem because most teachers in higher education have extensive knowledge in their respective disciplines, but not in pedagogical methods (Troncoso & Hawes, 2006). In his or her discipline, a teacher has a special language and practice, but less so in teaching, and this can lead to problematic interpretation of quality.

Moreover, much teaching is highly contextualized, so that teachers will interpret the same criteria in different ways solely because of the classroom, media, laboratory and other contexts in which they teach.

Specialist resources. It is common for a university to provide an office of instructional services, particularly for young and foreign speaking instructors. For example, at the Pontificia Universidad Católica de Valparaíso (Chile) is a center where raising the quality of student training and the quality of teaching are the most important purposes. To these ends, the center has different units working on evaluation of teaching, quality of teaching, development of new curricula and accreditation. Some of these services, such as training teachers, are offered to teachers who volunteer, but increasingly some programs have been assigning their younger teachers to it.

Some public universities of Mexico City have centers dedicated to evaluating and improving teaching, with specialists serving teachers on campus. In some institutions, such as the Universidad Autónoma Metropolitana (UAM) at Xochimilco, no specialist staffing exists to evaluate quality of teaching or to train teachers. The professors there are assessed informally by administrators and peers designated by the director of the appropriate academic division, and formally by a student questionnaire. Results have sometimes been used to award merit pay. For formative evaluation there, on certain occasions a few professors meet to discuss their problems. Occasionally, educational researchers meet to discuss the different educational practices at the university. Of course, individual instructors are engaged in some form of self-study, a few using teaching portfolios.

A hypothetical conversation between teachers

T1. I don't like them evaluating my teaching.

T2. Too intrusive?

T1. They don't understand the experiences I want my students to have.

T2. And you do it well?

T1. I could do better.

T2. Your students don't rate you highly?

T1. Ten percent hate it. They pull me down below average.

Formative evaluation of teaching. Evaluation is coming to know the quality of the act of teaching, finding merit and shortcoming in each part of teaching. The learning of students may or may not be an indicator of the quality of teaching (Cisneros-Cohernour, 1997). Formative evaluation is designed to help teachers improve their daily work and course fulfillment. As things stand, professors are the most important actors in this process because they alone are in place to observe and reflect upon their own practices (sometimes with the help of specialists or peers) and to decide the changes they will make (Arbesú, 2006). At times, formative evaluation will be supported by an administration looking for change to protect its autonomy and purposes. On most occasions, summative evaluation is what administrators prefer. Sometimes it is considered by both researchers and administrators that improving education is promoted through reflection on practice and that this reflection in turn generates knowledge about teaching practice (Brubacher, Case & Reagan, 2005; David, 2010, Rothstein, 2008).

Most researchers studying formative and comprehensive teacher evaluation are persuaded that it is useful (1) for changing methods of instruction, (2) for engaging students in upgrading the quality of classroom learning, (3) for promoting an aura of continuing professional education, and (4) for helping the academic community (teachers, students, administrative, authorities) understand the complexities of educational practices and in particular teaching (Cronbach, 1963). Unfortunately, however clear it is for professionals to realize and support these uses, the evidence is not sufficiently strong to persuade administrators to provide the resources needed for formative evaluation of teaching. The problems will be made clear in the following section examining the function of improvement at the university when given an evaluative data base.

Informally, everyone assesses teaching. Informal evaluation is constructed through perceptions and experiences of the people that participate in and observe the educational process: civil servants, teachers, students, administrators, etc. This type of evaluation occurs in all areas of the University and in all curricular and extracurricular activities.

Formally, assessment of teaching takes into account certain objectives and criteria a university uses to evaluate teaching quality. Educational researchers have assessed teaching for many years--but this field of knowledge is still under construction (Arbesú, Díaz Barriga, Elizalde, Luna, Rigo, Rueda & Torquemada, 2007; Ashcroft & Palacio, 2010;). Teaching assessment needs both formal and informal methods. Both help stakeholders refine insight into their accomplishments and problems.

Through this paper we have highlighted the complexity of education, of teaching, and therefore, the complexity of the evaluation. Both formal and informal assessment have been found useful for understanding the activities of teaching. In the next part we present strategies for moving from assessment to improvement of university functions. Highlighted again are the complexity of the university and the people within it.

The function of assessment in improving university operations

Advocates of assessment are optimistic that gathering data about university operations will lead to improvement. It sometimes does. But at this time there is no science of university management, no theory of higher education that has diagnostic quality. It is possible to learn much about how the university operates without understanding what could be done to improve it. From evaluation, for example, the Rector may learn that the Department of Forestry is captive to commercial interests but see no power or strategy for relief.

Figure 1 illustrated the responsibilities of formative evaluation of teaching, but did not show how evaluative knowledge can be converted into better teaching. And similarly for all functions of the university. Management of universities is greatly dependent on experience and intuition, however faulty those may be in some situations.

Formal assessment of most university functions, comprehensive and university wide, lies beyond their present capacities, just, as indicated above, as complete assessment of teaching is out of reach. But it can be improved. University functions traditionally are comprehensively assessed in informal ways; some should be made more formal. Informal ways are fallible, and formal assessment also is fallible. Validating and oversight responsibilities are needed for existing and expanding assessment mechanisms.

What good evaluation reveals is not only the quality of operations but their complexity. When it becomes accepted that simple relationships do not explain much, it is easier to resist simple solutions. Each operation has educational, financial, political, ecological, aesthetic and other complexities, foreground and background, and inquiry leads toward postulation as to what might be tried and toward search for examples of where it has been tried.

Mission statements are usually not much help because they tell what has been succeeding instead of what needs fixing. They are for aspiration and energizing more than guiding. They omit most of some problems and all of others. More realistic assessment of problems, issues and barriers is needed for better management. It is difficult to fix weaknesses in an atmosphere of self-promotion. Pressures for self-promotion have increased greatly, but good leaders will give more attention to fresh, skeptical and sustained reflection. Quality is not sufficiently defined in any mosaic of awards and ratings. The true worth of the university is much more mundane. Both aspirations and trajectories of optimism and skepticism are indicators of institutional health.

Strategies for formative evaluation. Four strategic ways are available for moving from assessment to actual improvement of university operations, including student instruction but extending through all functions. All four strategies are needed, but

probably not to be given equal priority. In each case there is need for attention to the heterogeneity of the university's stakeholders and the prospects of resources.

- The goal-based strategy. Those responsible for campus-wide planning and operations come to agreement as to what the university should be and prioritize shortcomings for remediation and new development as needed.
- The deficit strategy. Those responsible for campus-wide planning and operations come to agreement as to where the university is falling short of expectation, missing opportunities, and in some ways hurting its community, and prioritize the shortcomings to be worked on.
- The sector-study strategy. Increased support for recognition and alleviation of shortcoming is vested in departments, units, and other collectivities to come to agreement as to where their group is running smoothly, ineffective or dysfunctional.
- The self-study strategy. Increased support for recognition and alleviation of shortcoming at the level of the individual staff and faculty member is provided for self-study and coordination of improvement.

No matter how detailed the inquiry, the strategy for improvement will not come directly from indicators or ratings. Research on teaching may be suggestive; it will not be diagnostic. There is no alternative but to collect a wide array of informal and formal data for scrutiny and speculation as to where resources may best be spent. Often the individual staff member will have good guesses as to what changes in practice are feasible.

The whole and subdivisions of a university are patterned with uniqueness, with assets and contexts dissimilar. And universities differ as to the centrality of control. So it is not to be expected that evaluation and redevelopment will be the same or even visibly coordinated throughout the campus. As to the teaching function, the self-study model appears to have the greatest chance of success. Often the major function for central administrators is to demonstrate with conviction the need for assessment and to encourage sector leaders to devise their own strategies.

Evaluation of a university requires both formative and summative assessment, a recognition that the institution continues to mature and that there is a complexity of virtues and neglects that escapes easy measurement. The magnificent universities too have barriers to overcome. Assessment is not a guarantee for protection from disarray nor for quality control of growth, but comprehensive efforts to seek university quality can contribute to deliberative management. The greatest indicator of quality of a university might be the introspection of its administrators, staffs, and faculties.

Bibliography

- Ahire, S.L. (1997). Management science—Total Quality Management interface: An integrative framework, *Interfaces*, 27(6), 91-105. <https://doi.org/10.1287/inte.27.6.91>

- Ashcroft, K. and Palacio, D. (2010). *Researching into assessment and evaluation in colleges and universities*. London: Routledge.
- Arbesú, M. I. (2006). *La práctica de la docencia modular: El caso de la Unidad Xochimilco en la Universidad Autónoma Metropolitana*, México: UAM/Plaza y Valdés.
- Arbesú, M. I., Díaz Barriga, F., Elizalde, L., Luna, E., Rigo, M. A., Rueda, M. and Torquemada, A. D. (2007). La evaluación de la docencia universitaria en México: Un estado de conocimiento del periodo 1990 – 2004, en *Perspectiva Educativa*, 48, 27-58.
- Broudy, H.S. (1963). Historic exemplars of teaching method. In N. L. Gage, editor, *Handbook of Research on Teaching*. Chicago: Rand McNally.
- Brown, J. S. (1997). Common sense of purpose. In *What is a community of practice. Community Intelligence Labs*.
- Brubacher, J. W., Case, C. W. and Reagan, T. G. (2005). *Cómo ser un docente reflexivo. La construcción de una cultura de la indagación en las escuelas*, second edition. Madrid: Gedisa.
- Canales, A. (2003). Docencia: Tensiones estructurales en la valoración de la actividad. En M. Rueda, F. Díaz-Barriga, and M. Díaz (Eds.), *Evaluar para comprender y mejorar la docencia en la educación superior*. México: Universidad Autónoma Metropolitana-Universidad Nacional Autónoma de México-Universidad Autónoma de Baja California, 71-76.
- Cisneros-Cohernour, E. J. 1997. *Trade-offs: The use of student ratings results and its possible impact on instructional improvement*. University of Illinois: Unpublished report.
- Community Intelligence Labs (1997). *Communities of Practice*.
- Contreras, G. (2010). Diseño y operación de un sistema de evaluación del desempeño docente con fines formativos: la experiencia de la pontificia universidad Católica de Valparaíso, Chile. *Revista Iberoamericana de Evaluación Educativa*, 3(1e). http://www.rinace.net/riee/numeros/vol3-num1_e/art14.pdf
- Cronbach, L. J. (1963). Course improvement through evaluation. *Teachers College Record*, 64, 672-683.
- Cronbach, L. J. (1977). Remarks to the new Society. *Evaluation Research Society*, 1, 1.
- Cruz Souza, F. 1999. *Psicología Comunitaria*. Buenos Aires: Lumen/Hvanitas.
- Darling-Hammond, L. and Young, P. (2002). Defining “highly qualified teachers”: What does “scientifically-based research” actually tell us. *Educational Researcher*, 31(9), 13-25. <https://doi.org/10.3102/0013189X031009013>
- David, J. L. (2010). Using value-added measures to evaluate teachers. ASCD. <http://www.ASCD.org>.
- Dressel, P. L. (1971). Accreditation and Institutional Self-study. *North Central Association Quarterly*, 6, 277-87.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 4. <https://doi.org/10.1037/h0061470>
- Franke, A. H. (2008). New lessons dealing with sexual harassment. *Cronicle of Higher Education*. November 28.

- Lehman, I. J. (1975). Evaluating teaching. In Gephart, W. J., Ingle, R. B. and Saretsky, G., 1975. *The evaluation of teaching: National Symposium for Professors of Educational Research*. Bloomington, IN: Phi Delta Kappa.
- Rosu, L. M. and Stake, R. E. (2008). The assessment of teaching and learning in USA: Reflections on school quality and the role of teacher in public schools. *Romanian Pedagogical Journal*, 1(1), 49-66.
- Rothstein, J. (2008). *Student sorting and bias in value added estimation: Selection on observables and unobservables*. (Working Paper No. 170). Princeton, NJ: Princeton University and Cambridge, MA: National Bureau of Economic Research.
- Sadler R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, 30(2): 175-194. <https://doi.org/10.1080/0260293042000264262>
- Scriven, M. (1995). A unified theory approach to teacher evaluation. *Studies in Educational Evaluation*, 21, 111-129. [https://doi.org/10.1016/0191-491X\(95\)00009-J](https://doi.org/10.1016/0191-491X(95)00009-J)
- Stake, R.E. and Cisneros, E. J. (2000). Situational teaching evaluation of teaching on campus. *New Directions for Teaching and Learning*, 83, 51-72 <https://doi.org/10.1002/tl.8305>
- Stake, R. E. and Schwandt, T. A. (2006). On discerning quality in evaluation. In Shaw, I. F., Greene, J., & Mark, M., editors, *Handbook of Evaluation*. London: Sage. 404-418. <https://doi.org/10.4135/9781848608078.n18>
- Stufflebeam, D.L. (1981). Metaevaluation: Concepts, standards, and uses. In Berk, R. A., editor, *Educational Evaluation Methodology: The State of the Art*. Baltimore: Johns Hopkins University Press.
- Tinto, V. (1995). El abandono en los estudios superiores: Una nueva perspectiva de las causas del abandono v su tratamiento. En *Cuadernos de planeación universitaria*. 2 Época. Mexico: UNAM-ANUIES, año 6.
- Troncoso, K. and Hawes, G. (2006). A propósito de la evaluación por pares: La necesidad de semantizar la evaluación y las prácticas docentes. *Perspectiva Educativa*. 48, 59-73.
- Wenger, E. (1991). Communities of practice: Where learning happens. *Benchmarks*. Fall, 6-8.

Artículo concluido el 26 de julio de 2011

Stake, R.E., Arbesú, M.I., Contreras, G. (2017). Assessing the Quality of a University-Particularly its Teaching. *REDU. Revista de Docencia Universitaria*, 15(2), 125-141.

<https://doi.org/10.4995/redu.2017.6371>

Robert E. Stake

University of Illinois at Urbana – Champaign

Professor Emeritus, Educational Psychology

stake@illinois.edu

Professor Emeritus of Education, Educational Psychology, University of Illinois, 1998-present. Director of CIRCE, Bureau of Educational Research, University of Illinois, 1975-present. Professor Education, University of Illinois, 1965-1998. Associate Professor, Education, University of Illinois, 1963-1965. Assistant and Associate Professor, Educational Psychology, University of Nebraska, 1958-1963.

María Isabel Arbesú García

Universidad Autónoma Metropolitana-Unidad Xochimilco

Profesora e Investigadora de la División de Ciencias y Artes para el Diseño

isabel.arbesu@gmail.com

Especialización: Doctora en Educación. Categoría profesional: Profesor investigador titular “C” de tiempo completo. Estudios realizados: 2002, Doctora en Educación. Instituto de Ciencias de la Educación, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos. 1982, Licenciada en Ciencias de la Comunicación. UAM-Xochimilco, México. Líneas de investigación: Análisis y evaluación de la enseñanza universitaria. Representaciones sociales de los actores de la educación superior. Investigación sobre la docencia.

Gloria Contreras Pérez

Pontificia Universidad Católica de Valparaíso

Escuela de Pedagogía.

gloria.contreras@pucv.cl

Profesora de Matemáticas y Física por la Universidad de Playa Ancha, Chile, Magister en Evaluación Educativa Universidad de Playa Ancha, Chile, y Doctora en Ciencias de la Educación por la Pontificia Universidad Católica de Chile. Se especializa en evaluación de la docencia, tanto a nivel escolar como universitario, como también en evaluación a nivel de aula, en distintos niveles de escolaridad, con énfasis en su potencial formativo. En estas áreas se dedica especialmente a la investigación y formación de profesores, tanto de primaria como de secundaria, como también trabaja perfeccionando a docentes universitarios