



UNIUNEA EUROPEANĂ



GUVERNUL ROMÂNIEI  
MINISTERUL MUNCII, FAMILIEI  
ȘI PROTECȚIEI SOCIALE  
AMPOSORU



Fondul Social European  
POSDRU 2007-2013



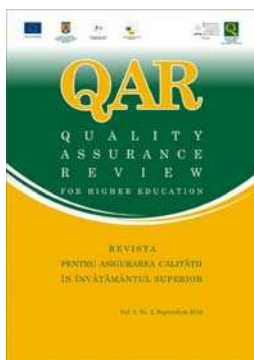
Instrumente Structurale  
2007-2013



MINISTERUL  
EDUCAȚIEI  
CERCETĂRII  
ȘI SPORTULUI



AGENȚIA ROMÂNĂ  
DE ASIGURARE A  
CALITĂȚII ÎN  
ÎNVĂȚĂMÂNTUL SUPERIOR



## Quality Assurance Review for Higher Education

### Quality Assessment in Higher Education: Research Public Universities in the United States

Maria Bucur-Deckard

Quality Assurance Review, Vol. 2, Nr. 2, Septembrie 2010, p. 102 – 108

**Publicat de:** Consiliul Agenției Române de Asigurare a Calității în Învățământul Superior - ARACIS

**Locul publicării:** București, România

**Tipul publicației:** tipărit, online

Quality Assurance Review este editată în cadrul proiectului „Asigurarea calității în învățământul superior din România în context european. Dezvoltarea managementului calității academice la nivel de sistem și instituțional”, Contract POSDRU/2/1.2/S/1, cod proiect 3933.

Pentru a reproduce un text din revistă este necesar acordul scris al colegiului de redacție al revistei Quality Assurance Review.

Răspunderea pentru textele scrise aparține autorilor.

Conținutul acestui material nu reprezintă în mod obligatoriu poziția oficială a Uniunii Europene sau a Guvernului României.

FONDUL SOCIAL EUROPEAN

Investește în  
**OAMENI**

# Quality Assessment in Higher Education: Research Public Universities in the United States

*Maria Bucur-Deckard*

*Indiana University Bloomington  
mbucur@indiana.edu*

**Abstract:** *This article aims to show the ways in which higher education in public universities in the United States with a ‘research one’ profile are moving towards a more intensive system of scrutiny in terms of both internal self-assessment of curriculum and faculty, as well as assessment by external funding agencies, be they private or governmental. Though arguably well intentioned and necessary, these different forms of quality assessment have had effects that are not always beneficial to the educational process, both in terms of improving quality, and also in terms of efficient use of time and resources. In short, though assessment of faculty is well established and has proven to be efficient in maintaining the quality of the research and teaching staff at research one universities, other forms of assessment, such as internal and external reviews, as well as periodic reporting to grant agencies, have not proven to be unassailably useful in understanding strengths and weaknesses in higher education, or proposing revisions to various programs.*

**Keywords:** *quality assessment, research one, quality improvement.*

This article aims to show the ways in which higher education in public universities in the United States with a ‘research one’ profile are moving towards a more intensive system of scrutiny in terms of both internal self-assessment of curriculum and faculty, as well as assessment by external funding agencies, be they private or governmental. Though arguably well intentioned and necessary, these different forms of quality assessment have had effects that are not always beneficial to the educational process, both in terms of improving quality, and also in terms of efficient use of time and resources. In short, though assessment of faculty is well established and has proven to be efficient in maintaining the quality of the research and teaching staff at research one universities, other forms of assessment, such as internal and external reviews, as well as periodic reporting to grant agencies, have not proven to be unassailably useful in understanding strengths and weaknesses in higher education, or proposing revisions to various programs.

## Definitions

Before launching into an analysis of an illustrative case study of how quality assessment does and does not work, I will set the stage by offering definitions of a few key concepts. To begin with, ‘quality assessment’ needs to be parsed into its two constitutive terms. ‘Assessment’ has in view several types of evaluation. Individuals engaged in the education process—faculty, staff, and students—periodically evaluate each other and themselves. This is both a contractual norm (obligatory annual faculty reviews, staff reviews, and student grades in individual classes) and also a habitual practice considered an important aspect of the teaching/learning process. Thus, even for informal workshops that are not repeated or evaluated externally, or in the case of one-time visiting lecturers, participants/students are generally asked to evaluate in some fashion the event they have taken part in. One can say there is a generalized acceptance, a culture of education assessment in institutions of higher education in the United States, be they private or public.<sup>1</sup>

---

<sup>1</sup> See for instance [http://www.msche.org/publications\\_view.asp?idPublicationType=5&txtPublicationType=Guidelines+for+Institutional+Improvement](http://www.msche.org/publications_view.asp?idPublicationType=5&txtPublicationType=Guidelines+for+Institutional+Improvement), and all publications attached to the page (accessed 5 July 2010).

There is, however, less agreement upon how to assess and what specific types of assessment can in fact tell us about the information being evaluated.<sup>2</sup> Thus, ‘quality’ is a contested term, both with regard to the assessment itself, and also in reference to that which is being evaluated (let’s call it ‘education’ as an overarching umbrella to cover all types of learning/teaching experiences in a university). For example, educators disagree over the meaning of a syllabus in terms of its promise for a quality learning experience in the classroom. Some faculty believe that detail and precision are guarantees of a well-thought, clear, and engaging learning experience, while others prefer a briefer version as a means to stimulate students’ imagination and preserve a level of mystery and freedom in the teaching/learning process throughout the semester. Pedagogical research is also divided over the effectiveness of either approach, since the classroom experience is ultimately live performance and depends a great deal both on the personality and skills of the instructor, as well as the personality of the class (individually and collectively).

The term ‘research one’, invoked at the beginning of this paper, refers to top-tier research universities in the United States.<sup>3</sup> The criteria for establishing one’s standing are set by a consortium of research universities, private and public, and combine: standards of admission (e.g., SAT scores, GPA averages, extracurricular activities of students enrolled, especially with regard to academic, public service, and professional development, national academic awards/fellowships garnered by the entering students, and more recently knowledge of foreign languages and experience living abroad); students’ grades and ratio of graduation, as well as their placement into post-graduate programs at other research one universities; the faculty’s standing in their respective fields in terms of research (as defined by each field); and the types and amount of competitive grants/awards garnered by its faculty and students. These standards and the annual results of evaluations by *Newsweek* and a few other independent organizations are published in various forms; they constitute an important basis for student and faculty recruitment, as they are considered by both constituencies as impartial and reflective of the quality of research and training available at the universities surveyed.<sup>4</sup> There are critics of these measurements, but no significant change in the parameters of evaluation has taken place over the past five years.

Though similar discussions about what it means to provide ‘quality assessment’ of higher education take place on all campuses across the United States, there are also important differences among the contexts in which they operate – private research universities, public research universities, private liberal arts colleges, public community colleges, etc. For the purposes of this essay, I will focus solely on the public research university context, using Indiana University, the Bloomington campus (IUB) as the main case study. Briefly, a public research university attempts to combine two fundamental goals: (1) a commitment to accessibility for a wide variety of students, both in terms of their cultural/economic background, and also in terms of their academic abilities, as demonstrated by grades in high school and scores on national tests (especially SAT); and (2) a desire to be a leader in scholarly research across a variety of fields, from the hard sciences to the humanities and professional schools (from business to law and public administration).<sup>5</sup> Successfully combining the two goals is difficult, as accessibility for students from a broad spectrum of educational backgrounds often comes at the expense of maintaining high admission and graduation standards, which are an essential measurement of a university’s standing among its research one peers.

Public universities are also defined by important limitations in how they govern themselves, a very important element in terms of overall assessment and self-assessment. Because they receive important tax benefits and significant funding (though decreasingly so in the past decade) from the

<sup>2</sup> Mark N.K. Saunders and Susan M. Davis, “The Use of Assessment Criteria to Ensure Consistency of Marking: Some Implications for Good Practice”, *Quality Assurance in Education*, vol. 6, no. 3 (1998): 162 – 171; Sandra Mathison and E. Wayne Ross, eds., *Defending Public Schools. Vol. IV. The Nature and Limits of Standards-Based Reform and Assessment* (Westport, CT, Praeger: 2004), available at <http://www.questia.com/PM.qst?a=o&d=106679962> (accessed 5 July 2010). *N*

<sup>3</sup> See <http://math.la.asu.edu/~kuang/ResearchI.html> (accessed 5 July 2010).

<sup>4</sup> See <http://bestcollegerankings.org/popular-rankings/> (accessed 5 July 2010).

<sup>5</sup> See <http://www.iub.edu/provost/campus/mission.shtml> (accessed 5 July 2010).

state in which they reside, they are beholden to the politics of that state in various ways. At IUB, this relationship with state politics translates in two significant ways: the state of Indiana has an appropriations committee, made up of politicians elected every two years by citizens residing in the state, which determines both the amount they are willing to give the university, and also the conditions under which the money is to be spent. Such conditions are sometimes connected to enhancing the economic development in the state<sup>6</sup> and generally privilege short-term return and professional schools. This context precludes a disinterested and academically-driven assessment of funding being dispersed, interfering with the autonomy of the university to focus on less immediate and concrete goals such as creating jobs in the state.

The appropriations committee also sets up the increase in tuition rate that is acceptable to the state.<sup>7</sup> Thus, even if the amount appropriated by the state is going down significantly from one year to the next (as was the case in the last two years), the university is not free to try and make up that deficit from increasing the tuition. This fact indirectly but very significantly shapes the fund-raising agenda of the university in two ways significant for quality assessment of the educational process: (1) faculty are increasingly pressured to apply to various agencies, private and public, to finance various educational projects, with the result of moving one's focus in education according to where the money comes from and under what conditions; and (2) a larger portion of faculty time is consequently devoted to fund-raising, which means that either research or teaching suffer.

Finally, and probably most importantly in terms of the overall ability of a state university to govern itself and set up an overarching agenda for pursuing academic excellence, the nine-member board of trustees, which ultimately approves all decisions – curricular, administrative, or financial, including internal and external system-wide assessments – of the university, is appointed by the governor of the state, with three exceptions, the members elected directly by the alumni.<sup>8</sup> Thus, the leadership of the university is overall in the hands of the top elected official of the state. Of all constituencies who participate in academic life of the university, until 1 July 2010, faculty have had no representation on the board of trustees.

Whether private or public, universities perform and are subjected to quality assessment on two general fronts: (1) the quality of the curriculum delivered to students; and (2) the quality of the faculty performing these services. There are differences in how these processes take place at a research one and, respectively, public university. In what follows, I will concentrate on general standards accepted across the United States, but will also draw attention to the particularities of the context of my case study.

## Quality of Curriculum Assessment

Until the past two decades, the idea of a standardized 'general curriculum' was not a point of much discussion on most university campuses in the United States.<sup>9</sup> However, since the 1990s the proliferation of increasingly diverse institutions of higher education, some highly specialized, others overtly general, all trying to attract a paying base of students, private donors, and federal grants, has driven the higher education administration and faculty towards trying to justify their work in terms of broad social ideals and utility. Though humanists and social scientists have their own proclivities

<sup>6</sup> See for instance IUB's current statement on relations with the State appropriations committee at <http://www.gov.indiana.edu/state/appropriations/index.shtml> (accessed 5 July 2010).

<sup>7</sup> <http://www.iu.edu/~upira/reports/standard/doc/fact%20book/fbook01/aprfee01.html> (accessed 5 July 2010).

<sup>8</sup> <http://newsinfo.iu.edu/news/page/normal/14863.html> (accessed 5 July 2010). This year's elections/appointments introduced one new important constituency among the board's members: a faculty member, Bruce Cole, whose credentials (for both academics and also the Republican Governor) were enhanced through his chairmanship of the National Endowment for the Humanities under George W Bush. The only alumni seat open for elections was won by a woman, who remains one of two female members on the nine-member body. She ran for the seat against five other candidates, all male.

<sup>9</sup> For a recent essay on the issue of general education standards in institutions of higher education in the United States, see Louis Menand, *The Marketplace of Ideas: Reform and Resistance in the American University* (New York: W. W. Norton & Company, 2010); see also Peter Salins' essay, "How Universities Got This Way", *Minding the Campus*, [http://www.mindingthecampus.com/originals/2010/02/review\\_essay\\_by\\_peter\\_d.html](http://www.mindingthecampus.com/originals/2010/02/review_essay_by_peter_d.html) (accessed 5 July 2010).

for theorizing around the notion of critical thinking and socially meaningful life-long skills, they have been less able to craft a coherent and persuasive discourse than professional schools, which have been more successful in describing graspable goals, both in terms of specific professional skills, such as public policy analysis or IT programming, and also in terms of financial success. In fact, the humanities in particular have been in a state of discursive disarray and inability to articulate a strong voice in academia.

With the reshaping of the learning environment towards a more transnational and virtual environment, what it means to have socially meaningful and useful skills has also become more complex and in need to articulate in relation to other more traditional goals of cultural literacy and critical thinking. In response to these pressures from education specialists, as well as the economy and students themselves, many universities across the United States today are developing new types of general education curricula. Instead of a handful of core courses in English, math, social sciences, and a few odd choices in the hard/natural sciences, faculty are coming together to create other clusters of skills. At IUB this has meant a healthy and at times heated debate over the past six years, with the result of new general education requirements coming into place in the fall of 2011, which will add to the mathematical and literature basic skills such new clusters as world cultures, public service, and cultural diversity.<sup>10</sup> A cluster focusing on sustainability may also join the general education curriculum sometime in the future.

The new requirements came about together with a set of assessment measures, to ensure that existing courses would count towards only if their content and approach could be demonstrated to fit the explicit intellectual goals of the new curriculum—above all, enabling students to think critically and at the same time appreciatively about the diverse world in which they live, and to navigate its complexities in ways that are both sustainable (financially and socially) and also personally fulfilling. Faculty may choose to continue teaching the same courses as before, but such courses will not count towards a growing number of core general education courses unless vetted according to the goals described above by a multidisciplinary committee of peers. The incentive for reforming new courses will be enrolments, which presumably would become an important financial resource for individual units on campus in negotiating for further graduate student funding, new hire lines, or merit-based salary increases. In truth, however, faculty with tenure may in fact choose to disregard such incentives and concentrate on their own personal goals, regardless of the costs for their academic units—they may choose to focus on research and not innovative teaching, on writing grants and not new courses, as part of their understanding of what a research one university means.

Another potential problem with the new curriculum is the ways in which it can easily slip down a different path, that has to do with a trend in governmental funding to demand quality assessment for any grants, without having in fact any clear and universally recognized standards. In other words, all university administrators are hearing more and more often that they need to show both how they spend government (and often also private) funds, and with what results. This is creating pressure on all academic units towards homogenizing the kinds of narratives and processes (research or teaching) they are performing, in order to persuade their administration to support them, and, more broadly, to garner the much sought after grants. An example in point should suffice as illustration. Biologists and language instructors are both trying to persuade often the same higher administration to provide financial support for their work in order to then apply for federal grants. The higher administration wants to see periodic assessment of how this financial support is spent. In one case, the money is going towards running large labs, with dozens of students and live animals performing experiments 24/7, with important findings sometimes noticeable over years of work, but with immediate results (the measurements taken regularly as an integral part of the scientific process) at observable and easy to report. In the other case, the 'subjects' are a diverse group of students learning difficult languages, on which the most expert research disagrees in terms of how many years it takes to move from

<sup>10</sup> <http://gened.iub.edu/> (accessed 5 July 2010).

beginner to ‘novice beginner’ to ‘intermediate’ to ‘intermediate advanced’.<sup>11</sup> Testing students regularly along various markers of skill acquisition is extremely costly, has to be done individually, and has proven thus far inconclusive, yet both administrations and donors are increasingly demanding that such assessments take place to the tune of often up to 1/3 of the grant. I am not aware whether any assessment of the efficiency of such significant spending has been attempted on a large scale, especially since the government can only agree, as donors, on the importance of learning difficult foreign languages such as Arabic, Farsi, Russian, and Mandarin, but not on what level of learning is required to serve the interests of the United States government (especially since such interests are generally very heterogeneous, from intelligence gathering to promoting US trade overseas).

Overall, the hard sciences are better poised at quality assessment as part of their discourse and methodology. The humanities and social sciences are slowly trying to grasp what it means to demonstrate progress in an academic setting in a measureable fashion from one semester to the next, but there are great difficulties in reaching a healthy compromise in agreeing between specific skills that can be taught in a fairly short time, and the notion that ‘critical thinking’ is a life-long process and that retention of skills and content is highly individualized process that no individual faculty actually controls.

## Faculty Assessment

Where all higher education institutions in the United States agree in terms of quality of assessment is the need to periodically evaluate individual faculty. The tenure system is not infallible in correctly assessing quality and seems to be somewhat under threat by the financial crisis of the moment. Yet all research universities and most liberal arts colleges are still very much beholden to the notion of periodic review of its tenure track and tenured faculty to ensure research and teaching standards. These standards are endorsed by the American Association for University Professors.<sup>12</sup> Briefly, most faculty at research universities are hired under the premise that within seven years (in fact the ‘trial period’ is five years, with the candidate assembling materials for tenure over the sixth year, and then awaiting the decision of the tenure committees over the seventh year) they will prove they are an important intellectual voice in their field/discipline, an excellent teacher, and a decent colleague in terms of university, departmental, and professional service overall.<sup>13</sup>

How these three aspects are assessed differs according to each discipline’s internally (and largely informally) agreed upon standards and according to each university’s own tenure/promotion stated criteria. But there are some basic types of assessment most institutions of higher education agree upon. First and foremost, a group of six to eight peers (all from a higher rank: full professors would evaluate associate professors and only senior associate professors would evaluate assistant professors) who are experts in the respective field and who work at peer institutions (research one for those being evaluated at research one institutions, etc.) are recruited to write evaluations of the candidate being assessed. The evaluators are generally recruited so as to not have any direct close academic and other links with the person being evaluated (e.g., dissertation advisor, co-author for a project or book/article, editor of a volume in which the candidate published a piece). Those asked to respond may in fact recuse themselves from this service to the profession and quite often do based on some or no explanation. The system works primarily on the presumption that those accepting to write the evaluation will be able to offer a detailed, expert, and impartial view of whether the colleague being evaluated is in fact someone who has made and will likely continue to make an important impact in

---

<sup>11</sup> See J. Charles Alderson and Jayanti Banerjee, “State-of-the-Art Review. Language Testing and Assessment”, *Language Teaching*, vol. 35, no. 2 (2002): 79-113; and Elana Shohamy, “Language Assessment and Program Evaluation. Performance Assessment in Language Testing”, *Annual Review of Applied Linguistics*, vol. 15 (1995): 188-211.

<sup>12</sup> <http://www.aaup.org/AAUP/pubsres/policydocs/contents/RIR.htm> (accessed 5 July 2010).

<sup>13</sup> The current standards at IUB, the case study discussed here, can be found at <http://www.indiana.edu/~bfc/docs/policies/tenurePromotionStatements.pdf> (accessed 5 July 2010).

the scholarly field of her/his expertise. These outside letters are in fact the most crucial piece of assembling the tenure/promotion file, as the people selected to write are not of the candidate's choice. She/he is sometimes asked to suggest names, but they are ultimately nominated by the department chair and tenure committee, and subsequently vetted by the appropriate dean.

In addition to these letters, the candidate submits a statement describing his/her activity over the period being evaluated, a type of intellectual autobiography. The peer-reviewed publications, grant narratives, and often manuscripts are brought in as evidence of excellence research activity, together with all awards and fellowships won by the candidate or her/his students. Course syllabi and student evaluations for each course taught along the five years, together with multiple letters of teaching evaluation from colleagues who have watched the candidate in the classroom are brought in as evidence of teaching abilities. Often students are also asked to write about their experiences being mentored by the candidate. Finally, service is assessed more in a quantitative manner, listing the number of committees and other types of service, such as organizing conferences, writing book reviews, or offering interviews in the press. Overall, research and teaching are the primary areas of assessment, with far less weight (and in no case I know of with the power to reverse the case) assigned to service. A faculty busy with the life of a public intellectual, for instance, and writing for *Newsweek*, *New York Times*, and *Saloon.com* will generally not be able to claim a standing in the profession (unless that profession is journalism) if she/he doesn't have peer-reviewed publications in his/her field of expertise and an excellent teaching record.

Change in one's record over the five years being evaluated is always an important element in assessment. Someone who starts with weak teaching evaluations and improves them over time, or who publishes a smattering of pieces in the first few years but then has a burst of peer-reviewed academic publications is generally viewed as having a normal path in the academic world. By contrast, a scholar with only one publication, with very few exceptions, to be usually argued for by the outside peer evaluators and not the home department or the candidate her/himself, is someone viewed as a risk and likely a weak scholar, and therefore may not be considered worthy of tenure.

The tenure case assembled, it is then brought before the home department by a group of peers from that unit, who present all aspects of the case in their own report and ask for the department to make a recommendation. Most often it is a positive recommendation, but there are many case where the positive recommendation (whether fair or not) hides doubts and offers veiled criticisms meant to raise the eyebrows of the evaluators further up the administrative ladder. After the vote in the department (recorded at such, with all numbers for, against, or abstaining now part of the record) the chair writes a letter with her/his own recommendation, which usually reflects the discussion in the departmental meeting. The dossier thus assembled is then sent to the dean's office, where a group of peers from the larger unit, across several disciplines, evaluates it and presents it to the dean. The dean has the freedom to accept that recommendation or to make her/his own decision, which has to be fully explicated if it differs from the decisions further down the line. If all committees and administrators up to this point agree that promotion should be granted, the case is then sent to the campus level, where the Provost/Chancellor, the President, and finally the Board of Trustees examine the case and make the final decision. If at any point one of these committees or administrators decides to deny promotion, the candidate is entitled to a contestation and, depending on the circumstances, may appeal to either the chair of the department or directly to the faculty senate appeals committee to adjudicate the matter. After the appeal is considered at the level where the rejection took place, the decision, whichever it may be, becomes final, and the faculty is either promoted or her/his contract ends at the end of the academic year.

There are similar means by which non-tenure faculty, as well faculty moving from associate to full professor are assessed. Likewise, external reviews of programs/units take place periodically, with a group of peer experts from other universities visiting the unit being reviewed (faculty, students, staff, administrators), reading materials being provided by the unit (and on occasion asking for additional materials), and offering recommendations for change, if necessary. These types of assessments are generally reliable, though mishaps and abuses do happen at various steps.

## Conclusions

Though scholars often disagree on what to assess and how in order to demonstrate the quality of their academic work, they do agree that peer reviewing is the single most important tool through which some quality assessment can be successfully (or rather reliably) implemented. The kind of periodic program-wide/system-wide assessment that is requested by funding agencies, be they the state, a federal institution, or a private foundation, is far more contested in terms of its reliability. Yet most public universities are feeling increasing pressure to seek such funding and comply with such assessment requests, especially as the cost of running a research one university is growing exponentially, while the revenue base is shrinking: state appropriations for higher education are going down, tuition is being kept by the same state appropriations authorities below the curve of covering budget increases, and federal grants are seeing more and more competitors and also louder choruses of politicians demanding ‘accountability’ without a clear definition of what that means. Spending more of their time trying to figure out the ‘right’ kind of assessment (meaning the kind that will please the right donor constituencies, and not necessarily and a type of assessment they actually consider reliable), faculty and administrators are having to spend less time either on research or on teaching activities, their main reason for being in academia (as opposed to professional fundraising, for instance). The quality of higher education in the United States thus hangs in the precarious balance between how much energy and time universities still want to allocate and want their faculty to spend on those primary two activities that makes us educators, versus courting donors and writing grants to suit politicians.

## References

- J. Charles Alderson and Jayanti Banerjee, “State-of-the-Art Review. Language Testing and Assessment”, *Language Teaching*, vol. 35, no. 2 (2002): 79-113;
- American Association of University Professors *Recommended Institutional Regulations on Academic Freedom & Tenure* <http://www.aaup.org/AAUP/pubsres/policydocs/contents/RIR.htm>;
- IUB *Tenure and promotion statements* <http://www.indiana.edu/~bfc/docs/policies/tenurePromotionStatements.pdf>;
- IUB Factbook, *State appropriation and student fee income over time*, <http://www.iu.edu/~upira/reports/standard/doc/fact%20book/fbook01/aprfee01.html>
- Sandra Mathison and E. Wayne Ross, eds., *Defending Public Schools. Vol. IV. The Nature and Limits of Standards-Based Reform and Assessment*, Westport, CT, Praeger, 2004;
- Louis Menand, *The Marketplace of Ideas: Reform and Resistance in the American University*, New York, W. W. Norton & Company, 2010;
- Peter Salins, “How Universities Got This Way”, *Minding the Campus*; [http://www.mindingthecampus.com/originals/2010/02/review\\_essay\\_by\\_peter\\_d.html](http://www.mindingthecampus.com/originals/2010/02/review_essay_by_peter_d.html) (accessed 5 July 2010).
- Mark N.K. Saunders and Susan M. Davis, "The Use of Assessment Criteria to ensure Consistency of Marking: Some Implications for Good Practice", *Quality Assurance in Education*, vol. 6, no. 3, 1998: 162 – 171;
- Elana Shohamy, “Language Assessment and Program Evaluation. Performance Assessment in Language Testing”, *Annual Review of Applied Linguistics*, vol. 15, 1995: 188-211;