

Quality Criteria for Self-Evaluation in Higher Education

Enrique Reboloso Pacheco

Baltasar Fernández-Ramírez

Pilar Cantón Andrés

University of Almería, Spain

ABSTRACT: The purposes of metaevaluation go beyond the traditional functions of accountability and enhancement. It helps guide strategic organizational change and legitimizes evaluation systems. Metaevaluation results can also be used to create checklists so that the persons responsible for any evaluation can revise, monitor, and control them by themselves. In this paper, the self-evaluation stage of an evaluation process in higher education is metaevaluated. The qualitative method applied analyses to the concerns and complaints that emerged during the work sessions. The results are structured around two key elements: the input (information and self-evaluation guideline) and the throughput (the committee and the process, including the meetings and dynamics of work). The information and self-evaluation guidelines are intended to be responsive, comprehensible, and reliable. The self-evaluation committee and process should be effective, efficient, and responsible with the task. The results are discussed with regard to theoretical proposals for evaluation quality that emphasize criteria such as responsiveness and technical quality of the evaluation.

KEYWORDS: *metaevaluation; evaluation quality; self-evaluation; higher education evaluation*

This research was funded by the former Spanish Ministry of Science and Technology, under Project Contract SEC2003-03230/PSCE.

In the 1980s, evaluation was an answer to a broad general change in education policies, summed up in a formula for decentralization in exchange for greater autonomy and self-management for universities. The Facilitator State defines the legal frames, transferring responsibility for growth, innovation, and diversification to the local institutions (Maasen, 1997; Neave & van Vught, 1991). Instead of traditional control systems, evaluation was called on as the preferred tool for quality assurance and accountability, as a counterbalance to increased centre autonomy ("the evaluative state," Neave, 1988). Import of

well-established quality management ideologies completed the extended discourse of evaluation as the main condition for institutional survival in the renewed context of European convergence (Haug & Tauch, 2001; Reboloso, 1999). European ministers and officials in charge of European Union members began a process intended to create a common European higher education area (*Bologna Declaration*, 1999). The change focused on the need for increasing legibility through a standardized structure of grades (bachelor-master degrees, common credit system), so as to increase mutual confidence through renewed evaluation and

accreditation systems. Main goals were to promote student and personnel mobility and to put European Higher Education in a good position for competing with a new and globalized offer of degrees (private universities, distance degrees, joint international degrees). Therefore, the Bologna process reinforced the role of institutional evaluation, notwithstanding the lack of criteria in the use of evaluation models and guidelines designed in the preceding years.

Evaluation models are anchored in both quality management and accreditation systems. The first emphasizes continuous improvement, strategic planning, personnel commitment, and dual orientation towards both clients and results (European Foundation for Quality Management [EFQM], 2001). The second requires self-evaluation systems aimed at improvement (Kells, 1983, 1995), varying the quality assurance guidelines by Kristoferssen, Sursock, and Westerheijden (1998). Evaluation assumes self-steering orientation. Any unit under scrutiny is capable of self-diagnosis, planning and implementing strategic improvement, and evaluating progress in five-year cycles. The procedure follows the four-stage model (van Vught & Westerheijden, 1994), in which the main tasks are self-study, external visit, external report, and improvement planning.

In the context of European convergence, responsibility, and improvement, the traditional purposes of evaluation have extended to both legitimating and strategic roles (Reboloso, Fernández-Ramírez, & Cantón, 2008). Strategic thinking leads universities to introduce innovations and reforms needed to be at the leading edge during this process. Legitimation demonstrates that the degrees offered are high-quality. One of the problems of current evaluation models and plans is that they were defined in a pre-Bologna context and have not been updated satisfactorily to cover the new purposes that are expected of them.

Goals of the Study

This study analyses quality criteria applicable to the self-evaluation stage and stresses the practical problems related to the structure and dynamics of working in committees. First we define the metaevaluation concept and describe some quality criteria applicable to the self-evaluation stage. In continuation, we go on to summarize the main results of qualitative monitoring of some self-evaluation committee meetings, attempting to point out their main worries, concerns, and complaints. Finally, a quality model is proposed for metaevaluation of self-evaluation, which is discussed with regard to some relevant general models. The model's applicability is emphasized and its use for the construction of future proposals arrived at through strategic thinking is suggested.

Concept and Criteria for Metaevaluation

Evaluation requires planning and implementing a systematic activity programme. Following its own general logic, the evaluation process and decisions can and must be systematically and critically assessed (metaevaluated), with the dual intention of demonstrating that the decisions were right and that findings lead to future improvement (Stufflebeam, 2001a). Useful quality criteria for this task are many and varied and depend on the intended goals, evaluation process characteristics, and researchers' conceptual and methodological orientations. An important question is the choice of metacriteria for deciding what specific criteria are of interest for the metaevaluation (Reboloso, Fernández-Ramírez, & Cantón, 2008).

Among the various methodological possibilities, an option is the preparation of checklists, that is, the selection of value criteria that indicate how the evaluation should be done in a given the evaluation context (Scriven, 2000). The criteria on the list come from

stakeholder interests and values and a theoretical framework with sufficient conceptual and empirical support (Leviton, 1994). Both are combined at the end of this paper, where we translate and synthesize into values the worries, concerns, and complaints of the self-evaluation committees participating in the study.

There is a considerable number of models and sets of criteria available for assessing evaluation quality, including the criteria of scientific validity (Chen, 1990), professional standards (Joint Committee, 1994), ethical principles (American Evaluation Association [AEA], 1995) and scientific research (Coryn, 2007; Coryn, Hattie, Scriven, & Hartmann, 2007). Chen tries to overcome the debate between a critical positivist paradigm and the suggestions of fourth-generation evaluation (Lincoln & Guba, 1986), by proposing four criteria for synthesis:

1. *Responsiveness*. Evaluation must be relevant and useful in producing the social improvement of those involved (for example, by improving timeliness, relevance, breadth of consequences, and equality in dealing with the interests of all of the stakeholders).
2. *Objectivity*. Evaluators must not have subjective biases that reduce the study's credibility (for example, checking to see that another evaluator could arrive at the same results under the same conditions or avoiding the influence of prior knowledge of the programme goals and interests).
3. *Confidence*. Confounding factors must be eliminated so stakeholders find the evaluation convincing or trustworthy.
4. *Generalizability*. Results of the evaluation can be applied to future circumstances or problems outside the limits of the specific programme.

Reboloso, Fernández-Ramírez, Cantón, and Pozo (2002) have adapted an evaluation

guideline to university administration services. They suggest three criteria applicable to the analysis and review of the contents or subjects of the guideline: a criterion of objectivity that combines the values of comprehensibility and clarity in the proper formulation of the guideline's subjects or sections; a criterion of responsiveness that inquires about the relevance and utility of these matters so the reflections suggested by the guidelines achieve a stronger impact on change and improvement in the management systems subject to evaluation; and a criterion of efficiency used to estimate the effort needed by the self-evaluation committee to discuss and respond to any questions in the guideline. These criteria do not represent a formal theoretical structure, but attempt to correctly formulate the contents of the guideline under an overall metacriterion of utility—which includes criteria of comprehensibility, effort, and relevance—for making evaluation guidelines that concentrate on the needs of their users and are more flexible, and, therefore, more useful.

Quality Criteria in Self-Evaluation

There are two basic requirements for starting self-evaluation: the information available for discussion and the evaluation guideline, which specifies the areas of interest the committees will discuss. Later, during the meetings, relevant factors have to do with the composition of the committee and work dynamics, including the attitudes of the participants toward the task. These four elements are essential to defining and analysing the self-evaluation stage (Reboloso, Fernández-Ramírez, & Cantón, 2003, 2008):

1. *Information mechanisms*. The various administration units that facilitate the information to the committees must be adequately coordinated (*data collection criterion*), the information must be appropriately processed and prepared

- (*information preparation* criterion), and there must be mechanisms that ensure their trustworthiness and objectivity (*data control* criterion) (Bustelo, 2003; European Association for Quality Assurance in Higher Education [ENQA], 2005; Joint Committee, 1994). Furthermore, all of the relevant information must be handed out to all the stakeholders in the unit evaluated (*distribution of information* criterion), emphasizing criteria such as free participation, open exposure and prior resolution of conflicts of interest among the stakeholders, transparency in prior promotion of the evaluation, and creation of a positive evaluation culture that encourages participation (AEA, 1995; van Vught & Westerheijden, 1994). This is the way to generate confidence among the participants and prevent rejection and value phobias that must be counterbalanced to ensure a viable, useful evaluation process.
2. *Evaluation guideline.* The guideline is an instrument for orienting discussion, avoiding committee efforts being distracted from their purpose of analysing, and improving their work units. The guideline should explain clearly what kind of evaluation is required of the participants (*judgement* criterion), and must teach them how to evaluate with clear, simple, exhaustive instructions on the goals and activities the committee is to perform (*content* criterion), in such a way that the work begins quickly and is not interrupted with questions about the procedure (International Association of University Presidents [IAUP] Commission, 2002; Reboloso, Fernández-Ramírez & Cantón, 2003). As mentioned above, Reboloso, Fernández-Ramírez, Cantón, and Pozo (2002) suggest three criteria for reviewing the quality of the subjects in evaluation guidelines: *objectivity* (comprehensibility of guideline expressions and contents), *responsiveness* (relevance of contents), and *effort* required to respond.
 3. *Committee.* All of the stakeholders should be represented on the committee (*representativeness* criterion), so their opinions can be taken into account if decisions are made that affect them (AEA, 1995; ENQA, 2005). All of the committee members should be fairly compensated for their efforts (*respect for personal equality* criterion) (AEA, 1995). Finally, there should be an appropriate number of members in the group to ensure its proper functioning, that is, enough to make discussions relevant and share the effort, but not so many as to impede meeting dynamics (*operability* criterion) (Reboloso, Fernández-Ramírez, & Cantón, 2003).
 4. *Process.* Active participation and collaboration of committee members is needed. Participation usually requires time to be spent in addition to an already overloaded agenda, so it is advisable for the activities to take a reasonable, and not an excessive, amount of time (*time management* criterion) (Bustelo, 2003). Moreover, everyone must be willing to discuss and come to a consensus (*democratic attitudes* criterion) (Reboloso, Fernández-Ramírez, & Cantón, 2003) and assume responsibility for the task (*personal obligations-execution of the task* criterion) (Stufflebeam, 2001b). Attitudes of apathy, “goldbricking,” delays or unconstructive criticism are unacceptable. Finally, the evaluation increases credibility and relevance when the stakeholders are involved in transferring their opinions and critical comments to the committees (*public participation* criterion) (Reboloso, Fernández-Ramírez, & Cantón, 2003).
- In conclusion, the original proposal for criteria suggests that self-evaluation quality requires having adequate information, before and during the meetings; a comprehensible,

relevant, and efficient guideline for evaluation; representative, operative, and fairly compensated self-evaluation committees; as well as a democratic, responsible climate for discussion in which everyone can contribute without a burdensome time commitment. The main values in play during self-evaluation fit in this narrow picture. Our working hypothesis is that they will also serve to approach the main concerns, complaints, and worries of the members of the committees in our study.

A Case Analysis: Self-Evaluation of Degrees at a Spanish University

Spanish universities have been making evaluations since the nineties, following the four-stage model (Mora, 2004; van Vught & Westerheijden, 1994). National and regional agencies (National Agency for Quality Assessment and Accreditation, Andalusian Agency of Evaluation of University Quality and Accreditation) promote coordinated plans for the purpose of evaluating all of the administration services and educational programmes about every five years. Part of the budget of every university is allocated annually for a certain number of evaluations to be conducted. Furthermore, administration services are contracting private accrediting agencies (Spanish Association for Standardisation and Certification IQNet, EFQM) for quality certification, and the logic of accreditation for implanting new study plans in the framework of the European Higher Education Space is becoming generalized. The variety of actions and evaluating agencies responds to the logic of multiple accreditation suggested by van Vught (1991) and van Vught and Westerheijden, 1994.

Contrary to an accreditation system, external reports on official evaluations are not binding; they are like friendly outside comments, which the in-house committee can take into consideration or not and make modifications or not in their final self-

evaluation report. Evaluation is completed with this final report, after being revised and corrected, after which the negotiation of plans for improvement begins with academic authorities (Unidad para la Calidad de las Universidades Andaluzas, 2004). That is why the self-evaluation stage is so important. It covers the main expectations and functions of evaluation, and it determines whether it will be consequential in system improvement and funding.

Study Method

The evaluators and a team of collaborators participated as observers in the self-evaluation committee meetings, collecting qualitative information on incidents and problems that came up during them. Exhaustive information was collected for two specific purposes: support while it was going on and, later, in defining a system of categories for future structured metaevaluation of institutional evaluations.

The original system of categories was derived from the quality criteria described above. Later, ad hoc modifications were made to classify observations, interrater reliability was calculated, and finally, the categories were reduced to a final set that was relevant, and meaningful.

The study method is qualitative, naturalist, and constructionist (Guba & Lincoln, 1989; Reboloso, Fernández-Ramírez, Cantón, & Pozo, 2000; Schwandt, 2002). Through intended nondefinition, the scant instructions given on the record sheets used for data gathering facilitated collection of qualitative complaints, concerns, and worries as they emerged in the participants' own language and in the heat of discussion, without bias toward observer beliefs. Without the imposition of a closed system of categories, the observers become tools for the interpretation of the situation, privileged participant-observers using their previous training and intuition as important factors in assisting in the selection of the data that should

be recorded (Harrison, MacGibbon, & Morton, 2001; Stronach, 2001). The flexibility gained with this type of participant observation provides an efficient way of recording events as they happen. The relevance and value of a subject of discussion is something that emerges from the dynamics of the meeting, observed in details that can be seen only from within, and advantage must be taken of having collaborators who join in the meetings without the limitations imposed by preset categories.

Participants

Six committees who were evaluating nine degrees in the University of Almería's Faculty of Humanities and Education Sciences participated in the study. Two of the committees did the self-evaluation for five degree specializations together, due to the high percentage of subjects in common among the different degrees and their three-year duration. It was also a way to reduce the effort required by the departments involved. Nevertheless, the committees made independent improvement plans for each different degree evaluated. The number of members on the committees varied from six to fourteen, with an average of nine. A representative of the administration staff participated on one of the committees, a student representative participated on three committees, and there were representatives of the faculty dean's office on all of the committees. That is, administration staff and students were not adequately represented. Members of the evaluation team participated in all of the sessions to assist in interpreting the guideline, make suggestions, take requests for information, and collect data while monitoring the process.

Instruments

The committees used the evaluation guideline edited by the Spanish Council of Universities (2002), and approved by the Andalusian

Regional Agency (UCUA, 2004). The guidelines included general instructions, a set of areas of interest that the committees should discuss to prepare their reports and plans for improvement, and a set of tables of quantitative indicators calculated and given to the committees by the responsible administrative units.

A series of questionnaires and record sheets were prepared for the parallel metaevaluation and monitoring activities. The observations noted on these record sheets are analysed and commented on further later in this article. The sheet identified the committee and the meeting and contained five open fields: the number of participants at the meeting, contents dealt with (subjects dealt with/subjects pending), problems interpreting the guidelines and solutions proposed, comments, and date of the next meeting (purpose, subject, and date). The observers note all the incidents they consider pertinent during the meeting in each field.

Procedure

The evaluation team was made up of two senior researchers and seven collaborators (Ph.D. students) who attended each of the self-evaluation meetings in pairs. All of them were previously trained in the contents of the guidelines and the evaluation process and received the same instructions for collecting data. One of the senior researchers participated in the first meetings and was responsible for organising and coordinating the process in order to identify and correct any problems the collaborators might have collecting data. The basic rule was to take unrestricted notes on any incidents that might slow down or impede the meetings, so the evaluation team would have direct information for the continued improvement of the process or its review in following years.

Difficulties in the Self-Evaluation Process

The self-evaluation process lasted from three to five months (mean = 16 weeks), with an average of 0.85 meetings per week. The six committees completed all the stages of the process. The improvement plans were approved and are currently underway. The average attendance of committee members at the meetings was 67 percent. Each meeting usually lasted slightly less than two hours. All the meeting places have the minimum favourable conditions (comfort, circular seating arrangement, and a flexible agenda for meetings).

The original system of categories was modified ad hoc to code 98% of the 603 comments collected during the self-evaluation meetings. The system was made up of seven categories (including “Others”). Reliability was estimated by three judges classifying the collected observations of one of the six committees. After the pertinent adjustments, an average of 90% agreement was considered satisfactory for system validation. The team of evaluators then discussed the observations in question until they were assigned to the most satisfactory category. Four observations were eliminated.

Each of the categories is discussed below. We use some literal examples of the data to illustrate our comments (the committee number and meeting are in parenthesis). The definitions and the number of cases in each category are shown in Tables 1 to 5. In spite of its numerical importance, the results in the “Degree Management” category are not discussed here, because they are not useful for analysing the quality or difficulties of the self-evaluation

process, which is the purpose of this article. For the same reason, the “Others” category, where the few comments that did not clearly fit in any of the remaining categories were assigned, is not mentioned either.

Quality Criteria for Self-Evaluation

In continuation, we synthesize and analyse the committees’ complaints and concerns (Tables 1 to 5) in order to suggest a list of values or quality criteria (see Table 6). It is a difficult task for interpretation (hermeneutic) and never foolproof. The result depends on the sensitivity of the observers who collect the data and our knowledge and skill in interpreting them. The same criticisms of the evaluation guideline may apply to our list of criteria, though difficulty for application and interpretation does not make them less useful for metaevaluation (Rossi, 1995; Shadish, Newman, Scheirer, & Wye, 1995). Values assist in understanding the experience of evaluation participants, their point of view on what works well and what does not; and guide the creation of checklists and work models to orient evaluation decisions and criteria used in metaevaluation efforts.

Information

Several university administration units prepared a file of quantitative data on the inputs, resources, and results of the degree evaluated. The comments on the quality of information with regard to needs, deficiencies, biases, or errors in information received are frequent at the first meetings. We distinguish four subcategories, as shown in Table 1.

Table 1
Comments on Category 1: Information

1. Information (125; 21%) Requirements, deficiencies, biases, or errors in information received by the committees at the beginning of the process for the purpose of undertaking discussion on each criterion and section in the guideline.
1.1. Insufficient data (39; 7%) Nonspecific criticisms on the lack of usefulness of the tables (“do not clarify much, little information, not very useful”. (C5, 6), and specific sections of the guideline in which there is not enough information. Not all the committees complain about the same problems, perhaps because they used their own information or preferred to answer intuitively. Consequently, the process is delayed and confidence in the results of discussions based on data that cannot be compared is reduced.
1.2. Biased or erroneous data (30; 5%) Data has specific errors or biases. For example, a survey delivered by the evaluators was criticised because it was hard to interpret and because of its lack of anonymity and representativeness (“many professors still haven’t answered” (C4, 6) or because they were supposedly not rigorous enough and lacking in validity (“they don’t think the survey is too valid because it is not anonymous” (C4, 8).
1.3. Demand for information (37; 6%) The committees demanded information of specific people or administrative units. In some cases, they suggested procedures for collecting information, especially questionnaires and surveys, and even “for discussion groups about the results” (C3, 2). Most of these demands were recorded during the first meetings.
1.4. Problems receiving information (19; 3%) The information did not get to the committee members. For example, they did not receive the tables or asked for the orientation summaries made by the evaluators to be sent to all of the members and not just to the committee chairperson.

Committees’ worries, concerns, and complaints concentrate exclusively on information received to assist them in the self-study discussions. No other type of prior information on the interests in play among the various stakeholders, the purposes of the evaluation, or success in creating an evaluation culture emerged from discussions (AEA, 1995; van Vught & Westerheijden, 1994). The committees focused on the immediate task of completing the self-report protocols, without criticisms or complaints about previous procedure being transferred to the record sheets. The underlying value criteria resulting in this first category are grouped in two blocks: responsiveness and technical quality of information (Chen, 1990; Reboloso, Fernández-Ramírez, Cantón, & Pozo, 2002). Evaluation is responsive if the information handed out to the self-evaluation committees

approximates its true requirements for performance of the task. Any information is not good enough. It has to be relevant to the discussions, that is, complete (necessary), suitable, and delivered at the right time for it to be used during the discussions (timeliness). The goal is to increase the usefulness of the information, which makes responsiveness the key criteria. The technical quality of the information is related to the confidence in data that are properly acquired, prepared and delivered (criteria of *preparation of information* and *data control*) (Bustelo, 2003; ENQA, 2005; JCS, 1988; Reboloso, Fernández-Ramírez, & Cantón, 2003). Specifically, confidence depends on clear, comprehensible results, without errors or bias; that is, they must faithfully represent the opinion of the faculty and other groups that have had the opportunity of participating in the process. Using conventional language, the data

are rigorous, objective, and reliable. In fourth-generation language (Guba & Lincoln, 1989), the data merit the confidence of their receivers, accepting the idea of intersubjective consensus as the criterion of validity (Lincoln & Guba, 1986).

Evaluation Guidelines

The evaluation guideline is a long document that reviews a multitude of subjects related to

degree management. It is no different from other similar guidelines and has the same difficulties, such as lack of examples, generic expressions, and the requirement for analyzing complex subjects without providing precise instructions of any kind. We distinguish two subcategories.

Table 2
Comments on Category 2: Evaluation Guideline

2. Evaluation Guideline (97; 16%) Criticism by the committees about the problems and difficulties associated with the various sections and tables in the evaluation guideline
2.1. Problems with the tables (26; 4%) The tables include poorly chosen, ambiguous information or concepts hard to understand, too generic or confusing. Interpretation is difficult and usefulness for discussion reduced. For example, “tables not very clear” (C1, 1), “hard to analyse” (C2, 1), or “the information in these tables does not help me prepare this point. Solution: Not use them” (C4, 9).
2.2. Problems with the guideline (71; 12%) There are problems understanding the guideline proposals because they are prone to confusing interpretation (“what to include in the degree trajectory milestones” (C4, 1), not knowing how to prepare the answer (“measurement of prestige—how do you quantify that?” (C4, 1), or what is asked of the committee is simply unclear (“what does <i>potential quality of the faculty</i> mean?” (C6, 10). “The guideline should be simpler” (C1, 14) and less technical, and the questions posed more realistic, pertinent and relevant in order to reflect the reality of the degree evaluated.

Committee concerns here also seem to concentrate on the immediacy of the daily task of self-study, with no criticism on the first part of the guideline, where the procedure and purpose of the work are explained to orient its use throughout the process (IAUP Commission, 2002; Reboloso, Fernández-Ramírez, & Cantón, 2003). The quality of the tables mainly depends on the comprehensibility, clarity, and responsiveness of the concepts that structure them. There are poorly chosen, ambiguous concepts and data other than those the committee expects, or too generic, providing overall information on the university or faculty, but not on the degree being evaluated. The guideline is not clear either, because it uses

terms that are confusing, too complicated, excessively general, or too specific, so the committees do not know exactly what is expected of them or what they should include in their answers. The task is not understood. The problems of responsiveness indicate that the guidelines are not adapted to the reality of the degree being evaluated, because the information available is insufficient or because the questions are unrealistic, scarcely pertinent or irrelevant to the self-evaluation. To summarize, the values implicit in the criticism of the guidelines are structured in the categories of responsiveness and comprehensibility already applied in the section above. The consequence of these

problems is difficulty in completing the self-study and loss of usefulness of the guidelines.

and the committee's solutions. We distinguish four subcategories, including one "Others."

Procedure

This category concentrates on internal organisation and work dynamics, its problems

Table 3
Comments on Category 3: Procedure

3. Procedure (123; 21%) Comments on the internal committee organisation and work dynamics, problems and solutions contributed by the members
3.1. Secretary (18; 3%) The secretary is a role accepted grudgingly, which causes arguments and rejection in three committees, because it fulfils a technical function with more responsibility, effort, and involvement (generally, taking notes on discussions, writing the self-report, making the corrections, and preparing the final versions). The solutions to the problem are varied, such as rotating the task, joining the chairperson and secretary functions, and even having nobody take on this function. Expressions such as "nobody wants to," "they fight," or "they argue" suggest a heated discussion in making the decision.
3.2. Report writing and revision (23; 4%) Report writing is usually done by preparing drafts, private reading by each member, and approval at later meetings. For example, after discussing some parts of it, the secretary prepares a draft, hands it out, and collects suggestions. Other solutions are to take a break in the meeting, for example, "thirty minutes for the secretary to write a summary" (C4, 12), or start by reviewing the reports from previous meetings.
3.3. Method of work (60; 10%) The organisation of work generates a variety of discussions beyond the question of the self-report. Some committees doubt whether to use the supporting material the evaluators have prepared to orient them in the interpretation of the guideline; the solutions proposed compare the two documents or even reject the material. The lack of congruence between one table and a section of the evaluation guideline led them to decide twice that "the tables are not going to be used ... for dealing with the guideline" (C5, 9). Some comments include decisions about the organisation of the following meeting or during the rest of the process, discussing the committee's responsibilities (C4, 2) or proposing that the meetings be recorded (C3, 4). Others deal with the type of data and information sources that should be used, for example, "they discuss whether the obvious perceptions each has from his own experience can be included in the report" (C1, 2). Finally, one of the committees discusses persistently whether they should and how to assume the responsibility of defining the goals of the degree as the guideline suggests.
3.4. Other (Procedures) (16; 3%) In this last group, we have included some practical procedural problems, such as difficulties in setting the time of the following meeting and problems derived from the composition of the committees that are to evaluate more than one degree.

The remaining categories (procedure, distractions, and participation) represent different aspects and problems of the dynamics of the self-evaluation sessions. Some of the criteria that are supposed to be important hardly appear, such as matters concerning the

composition of the committee or compensation received for the effort (AEA, 1995; ENQA, 2005; Reboloso, Fernández-Ramírez, & Cantón, 2003). The problems detected are interpreted from a general criterion of task clarity-comprehensibility and an efficiency

criterion. The participants require better prior training and practical instructions for decision making on the organization of work or distribution of functions, so that time is used for the self-study, and not in discussing how to do it. Organisation of the discussion, preparation of drafts and final versions of the self-report, or decisions about the time of the following meetings are related to the criteria of task clarity and comprehensibility (Reboloso, Fernández-Ramírez, Cantón, & Pozo, 2002). Punctuality or whether discussions are well-

focused and not dispersed are efficiency criteria linked to better time management (Bustelo, 2003).

Distractions

This category refers to discussions and delays that reduce the agility of the self-evaluation meetings.

Table 4
Comments on Category 4: Distractions

4. Distractions (71; 12%)

Discussions, digressions, and tardiness that reduce the agility of the self-evaluation meetings. The main reason for the delay is the appearance of discussions that have nothing to do with the guideline sections to be dealt with, for example, about other universities (C5, 3), other degrees (C5, 4), or digressing about the quality of the degree (C2, 1). Other delays are due to tardiness (C1, 12; C6, 4), repeated absence of some members ("many absences; only 2 members" C3, 8), telephone interruptions (C6, 7), and using too much time to review the previous results or anticipating later points of discussion (C4, 1). The problem of absenteeism was very important in one of the committees, even threatening continuity of the process (C5, 8).

This is a special case of the category above. The goal of improvement implicit in the results would be to make changes or controls to expedite the dynamics of the meetings. The main distractions are unnecessary discussion, tardiness, and interruptions due to outside factors. Some committee members are easily distracted from the discussion and wander off on subjects of no concern, increasing the rest of the members' perception of uselessness.

Tardiness is especially severe in some committees, as we have seen, almost threatening the continuity of the work. All of these distractions increase the time used to complete the tasks (time management criterion; Bustelo, 2003). The key criterion is efficiency. The committee chairperson must receive adequate training to direct the meetings and avoid losing time.

Table 5
Comments on Category 5: Participation

5. Participation (62; 10%)

This category includes deficiencies in the degree and quality of participation of the committee members. We distinguish three main types of comments:

a) Deficiencies in participation of committee members who “hardly contribute anything” (C3, 2) to the discussions, do not prepare for the scheduled task (“do not read what the guideline says” [C3, 2]) or who interrupt (“do not respect the rhythm of the group and interrupts with their chatter, on top of not bringing the material” [C5, 4]).

b) Complaints and situations that could have set off an incident due to the bad atmosphere in the group or disruptive participations.

c) Mentions about the good working atmosphere, with comments such as “fluid work, positive, enriching discussions” (C2, 8) or “there is a good atmosphere and work flows at a good rhythm” (C3, 16).

The remaining comments have to do with problems of the work overload produced by participation in more than one self-evaluation committee or combining two evaluations in just one committee, as well as the appearance of small groups talking during the meeting, producing nonorganized joint discussions.

Participation

The main hurdle to participation in improvement teams is overworked members. The sensation of losing time is a demotivating factor that should be avoided by imprinting the sessions with agility and productivity. The basic value of participation is responsibility, which each person assumes for the tasks that must be done, most especially, the person who takes on the function of secretary. We have found a series of complaints related to this point, synthesized in noncompliance with the tasks (not contributing to discussions, not preparing the task), lack of motivation and disruptive behaviour (interrupting the meeting, arguing). Compliance with such personal obligations is emphasized by Stufflebeam (2001b) as a requirement for the execution of the task, while Reboloso, Fernández-Ramírez, Cantón, and Pozo (2002) propose the effort required as an indicator of efficiency.

criteria associated with each of them (Reboloso, Fernández-Ramírez, & Cantón, 2003; Reboloso, Fernández-Ramírez, Cantón, & Salvador, 2005). The classification of the incidents and observations recorded during the process produced a system of seven categories. We have concentrated our efforts in the interpretation of those five categories directly dealing with specific self-evaluation problems.

The final model contextualizes and summarizes the quality criteria available for metaevaluating a self-evaluation process (Table 6). The model shapes the values displayed during the self-evaluation meetings in a common university degree evaluation system. It can be adapted in a metaevaluation questionnaire or used as checklist for evaluation.

Conclusion

We start out with a set of four key elements in the self-evaluation process and a set of quality

Table 6
Elements and Quality Criteria for Self-Evaluation

Key elements	Areas of incidents	Quality criteria
Information	Quality, reception, and demand of information	Responsiveness: usefulness, relevance, complete information, suitable information, timeliness Technical quality: objectivity, clarity, comprehensibility, reliability, correctness, unbiased, confidence-worthy
Evaluation guideline	Tables and sections of the guideline	Responsiveness: relevance, adjusted concepts Quality of concepts: comprehensibility, clarity, ambiguity, complexity, focus
Process	Procedure, distractions	Comprehensibility Effort, involvement Efficiency: punctuality, time management
Committee	Participation	Responsibility, motivation Efficacy, workload, completion of tasks Efficiency

Two subsets of criteria can be observed in Table 6, depending on whether we talk about inputs (information, evaluation guideline) or the process itself (process, committee). The evaluation guideline and the data that are handed out to the committees are the basic resources required for the process to function. Responsiveness and technical quality are key values (Chelimsky, 1983). The data must be suitable for the needs of the committee and assure certain technical characteristics to gain their confidence. The introduction of the confidence element is a step beyond the conventional criteria of scientific quality, which do not consider the opinions of the persons the information is for as a matter of validity (Chen, 1990). Confidence brings us closer to the issues of responsiveness and utility—characteristics of applied research, social intervention, and management, which distinguish evaluation as an activity different from traditional laboratory research.

The process and committee elements shared some conventional perspectives of efficacy and efficiency, also characteristic of action contexts. The process emphasizes efficiency by valuing time and through the emerging problems of

distraction. The committees must have persons on them who are able to work well and are motivated to assume responsibilities. Even more so, both elements lead to a discourse in which people stand out as a key factor in the self-evaluation task. The quality of the process is not defined by the results (the evaluation report the committee is preparing), but by performing the task rapidly and correctly. The role of evaluator has been delegated to the members, who have very limited training in the subject and very little time to lose. Practical solutions are welcome and technical criteria are complemented by the required motivation of the members to assume responsibilities.

This set of values is coherent with known evaluation quality criteria proposals, such as the criteria of usefulness and technical adequacy (Chelimsky, 1983; Stufflebeam, 1974), responsiveness and objectivity (Chen, 1990), or accuracy and utility (Joint Committee, 1994). However, we miss an ethical perspective among the worries and concerns of the committee members, like the AEA (1995) ethical principles, only visible in matters of motivation, involvement, and responsibility for taking on self-study tasks.

Limitations of the study are clear. It needs to be extended to include all the stages in the entire evaluation process, from previous political decisions and logistics, to the postevaluation process of implementation and evaluation of the improvement plans. Jeliazskova (2002) and Lawrenz, Keiser, and Lavoie (2003) discuss how to improve the external evaluation stage. Reboloso, Fernández-Ramírez, and Cantón (2008) analyse a panoramic scope of the whole process. Many of the checklists by the Western Michigan University Evaluation Center Checklist Project (www.wmich.edu/evalctr/checklists) also suggest criteria that can be directly applied to the metaevaluation of higher education degree evaluations.

The study emphasizes a problem-based perspective, as the main purpose is to find the problems in the self-evaluation stage to support and enhance the evaluation system in the years to come. For example, comments on positive topics could be ignored by the observers or simply categorized as distractions from the discussions. Therefore, results should be biased in a problem-focused discourse, leaving other values aside.

The values proposed in the model are the results of a first effort in interpretation. From the logic of verification, discussions with stakeholders in higher education should be the next stage in confirming the validity of the model, following fourth-generation evaluation through the iterative stages of discussion and negotiation with stakeholders (Guba & Lincoln, 1989). Thinking strategically (Mintzberg, 1993), these values would be used as an input resource to negotiate the values to be maximized in future improved evaluation processes (Reboloso, Fernández-Ramírez, & Cantón, 2008). Current institutional evaluation in higher education fails to transmit to the participants an attitude of flexibility and freedom to interpret the self-evaluation guideline proposals in the way most relevant to maximize their goals. Institutional management has no correct or

wrong answers, as measured by a questionnaire in which evaluation participants could value their work on a ten-point scale. Evaluation guides are not questionnaires, nor instruments seeking any scientific truth, but guides for reflexion on and construction of an institutional future. Self-evaluation committees should be better trained in order to understand the guideline contents and its possibilities, and they should trust more in their intuition as participants to appropriate the ideas in the guideline and take advantage of them for their own strategic proposals for the future.

References

- American Evaluation Association. (1995). Guiding principles for evaluators. In W. R. Shadish, D. L. Newman, M. A. Scheirer, & C. Wye (Eds.), *Developing the guiding principles. New Directions for Program Evaluation*, 66, 19-26.
- Bologna Declaration of 19 June 1999: Joint declaration of the European Ministers of Education convened in Bologna. (1999). Retrieved November 6, 2008, from http://www.ond.vlaanderen.be/hogeronderwijs/bologna/documents/MDC/BOLOGNA_DECLARATION1.pdf
- Bustelo, M. (2003, October). *Metaevaluation as a tool for the improvement and development of the evaluation function in public administrations*. Paper presented at the 5th Biennial Conference of the European Evaluation Society. Seville, Spain.
- Chelimsky, E. (1983). The definition and measurement of evaluation quality as a management tool. In R. G. St Pierre (Ed.), *Management and organization of program evaluation. New Directions for Program Evaluation*, 18, 113-126.
- Chen, H. T. (1990). *Theory-driven evaluations*. Newbury Park, CA: Sage.
- Coryn, C. L. S. (2007). The 'holy trinity' of methodological rigor: A skeptical view.

- Journal of MultiDisciplinary Evaluation*, 4(7), 26-31.
- Coryn, C. L. S., Hattie, J. A., Scriven, M., & Hartmann, D. J. (2007). Models and mechanisms for evaluating government-funded research: An international comparison. *American Journal of Evaluation*, 28(4), 437-457.
- European Foundation for Quality Management. (2001). *EFQM Excellence Model: Public sector and volunteer organisations*. Madrid: Club Gestión de la Calidad.
- European Association for Quality Assurance in Higher Education. (2005). *Standards and guidelines for quality assurance in the European Higher Education Area*. Retrieved November 6, 2008, from <http://www.enqa.eu/files/ENQA%20Berg%20Report.pdf>
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth-generation evaluation*. Newbury Park, CA: Sage.
- Harrison, J., MacGibbon, L., & Morton, M. (2001). Regimes of trustworthiness in qualitative research: The rigors of reciprocity. *Qualitative Inquiry*, 7(3), 323-345.
- Haug, G., & Tauch, C. (2001). *Towards the European Higher Education Area: Survey of main reforms from Bologna to Prague: Summary and conclusions*. Retrieved November 6, 2008, from http://www.ond.vlaanderen.be/hogeronderwijs/bologna/documents/EUA_Trends_Reports/TRENDS_II-April2001.pdf
- International Association of University Presidents. (2002). *Towards a worldwide quality register for quality assurance and accreditation agencies*. Retrieved November 6, 2008, from http://www.iaups.org/documents.php?cat_id=1
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards*. Thousand Oaks, CA: Sage.
- Jeliazskova, M. (2002). Running the maze: Interpreting external review recommendations. *Quality in Higher Education*, 8(1), 89-96.
- Kells, H. R. (1983). Improving institutional performance through self-study. In K. E. Young, C. N. Chambers, H. R. Kells, & Associates (Eds.), *Understanding accreditation* (pp. 119-132). San Francisco: Jossey-Bass.
- Kells, H. R. (1995). Building a national evaluation system for higher education: Lessons from diverse settings. *Higher Education in Europe*, 20(1-2), 18-26.
- Kristoferssen, D., Sursack, A., & Westerheijden, D. F. (1998). *Manual of quality assurance: Procedures and practices*. Brussels: EC/Phare/ETF.
- Lawrenz, F., Keiser, N., & Lavoie, B. (2003). Evaluative site visits: A methodological review. *American Journal of Evaluation*, 24(3), 341-352.
- Leviton, L. (1994). Program theory and evaluation theory in community-based program. *Evaluation Practice*, 15(1), 89-92.
- Lincoln, Y. S., & Guba, E. G. (1986). But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. In D. D. Williams (Ed.), *Naturalistic evaluation. New Directions for Program Evaluation*, 30, 73-84.
- Maasen, P. A. M. (1997). Quality in European higher education: Recent trends and historical roots. *European Journal of Education*, 32(2), 111-127.
- Mintzberg, H. (1993). *The rise and fall of strategic planning: Reconceiving roles for planning, plans, planners*. New York: The Free Press.
- Mora, J. G. (2004). A decade of quality assurance in Spanish universities. In S. Schwarz, & D. F. Westerheijden (Eds.), *Accreditation and evaluation in the European higher education area* (pp. 421-443). Dordrecht, The Netherlands: Kluwer.
- Neave, G. (1988). On the cultivation of quality, efficiency and enterprise: An overview of recent trends in higher education in Western Europe 1986-1988. *European Journal of Education*, 25(2-3), 273-283.

- Neave, G., & van Vught, F. (1991). *Prometheus bound: The changing relationship between government and higher education in Western Europe*. Oxford, NY: Pergamon.
- Reboloso, E. (1999). *La evaluación de la calidad como estrategia de supervivencia and futuro de la universidad*. [Evaluation of quality as a strategy for the survival and the future of the university.] Almería, Spain: Universidad de Almería.
- Reboloso, E., Fernández-Ramírez, B., & Cantón, P. (2003). *Guía de evaluación de servicios universitarios*. [Guide for the evaluation of university administration services.] Almería, Spain: UCUA.
- Reboloso, E., Fernández-Ramírez, B., & Cantón, P. (2008). Responsibility of educational institutions for strategic change. *Journal of MultiDisciplinary Evaluation*, 5(10), 5-20.
- Reboloso, E., Fernández-Ramírez, B., Cantón, P., & Pozo, C. (2000). El papel de la investigación cualitativa en la evaluación de los servicios universitarios. [The role of qualitative research in the evaluation of university administration services.] *Cuadernos IRC*, 4, 65-82.
- Reboloso, E., Fernández-Ramírez, B., Cantón, P., & Pozo, C. (2002). Metaevaluation of a total quality management evaluation system. *Psychology in Spain*, 6(1), 12-25.
- Reboloso, E., Fernández-Ramírez, B., Cantón, P., & Salvador, C. M. (2005). *Criterios de calidad de las evaluaciones universitarias*. [Criteria of quality of university evaluations.] Paper presented at the IV Conferencia de la Sociedad Española de Evaluación. Retrieved November 6, 2008, from <http://www.sociedadevaluacion.org/docs/conferencias/4thMadrid2005/com122.pdf>
- Rossi, P. H. (1995). Doing good and getting it right. In W. R. Shadish, D. L. Newman, M. A. Scheirer, & C. Wye (Eds.), *Guiding principles for evaluators*. *New Directions for Program Evaluation*, 66, 55-59.
- Schwandt, T. A. (2002). *Evaluation practice reconsidered*. New York: Peter Lang.
- Scriven, M. (2000). *The logic and methodology of checklists*. Retrieved October 6, 2008, from http://www.wmich.edu/evalctr/checklists/papers/logic&methodology_oct05.pdf
- Shadish, W. R., Newman, D. L., Scheirer, M. A., & Wye, C. (Eds.). (1995). *Developing the guiding principles*. *New Directions for Program Evaluation*, 66.
- Spanish Council of Universities (2002). *Guía de Evaluación de la Titulación, II Plan de la Calidad de las Universidades*. [Guide of evaluation of degree, II plan of quality of universities.] Madrid: Secretaría General del Consejo de Universidades.
- Stronach, I. (2001). The changing face of responsive evaluation: A postmodern rejoinder. *New Directions for Evaluation*, 92, 59-72.
- Stufflebeam, D. L. (1974). *Metaevaluation*. Occasional Paper Series #3. Kalamazoo: Western Michigan University Evaluation Center.
- Stufflebeam, D. L. (2001a). The metaevaluation imperative. *American Journal of Evaluation*, 22(2), 183-209.
- Stufflebeam, D. L. (2001b). Evaluation checklists: Practical tools for guiding and judging evaluations. *American Journal of Evaluation*, 22(1), 71-79.
- Unidad para la Calidad de las Universidades Andaluzas. (2004). *Guía de autoevaluación de titulaciones*. [Guide for the self-evaluation of grades.] Córdoba, Spain: Author.
- van Vught, F. A. (1991). Higher education quality assessment in Europe: The next step. *CREaction*, 96(4), 61-83.
- van Vught, F. A., & Westerheijden, D. F. (1994). Towards a general model of quality Assessment in higher education. *Higher Education*, 28(3), 355-371.