A Generic Performance Evaluation Model for Colleges and Universities

Revised August 2010
Department of Computer Science, Keene State College, NH

Elvis C. Foster is Assistant Professor of Computer Science at Keene State College, New Hampshire. He holds a PhD in Computer Science from University of the West Indies, Mona, Jamaica, and is a PhD student of Organizational Management at Capella University, Minneapolis, Minnesota.

Abstract

This paper examines the problem of teacher performance evaluation in colleges and universities. It commences by examining existing literature on the subject. It then highlights problems with existing peer-review models by examining what various institutions have done, followed by two case studies. It then proposes a performance evaluation model that may be applied to any similar institution.

1. OVERVIEW

Teacher performance evaluation has the become a very topical issue over the past two decades, with strong support for the American Federation of Teachers (AFT) and the National Education Association (NEA), two leading teacher associations (Murray and Grant, 1998). A significant component of teacher performance evaluation is peer evaluation. However, peer evaluation has posed challenges to colleges and universities for some time. It has been widely accepted that teacher peer evaluation is a useful exercise for colleges and universities (Braskamp and Ory, 1994; Weimer, 1990). Yet, research indicates that in its current practice, teacher peer evaluation is not the best indicator of teacher effectiveness (Yon, Burnap and Kohut, 2002; Odden, 2004). Research also indicates that teacher peer evaluation as currently practiced, is too subjective, and therefore ineffective in meeting the very objective it sets out to achieve — that of improving teaching effectiveness (Alicias, 2005; Cohen, 2003; Peterson, 2004). Moreover, it has been argued that teacher peer evaluation does not help the teacher (Bushman, 2006). Finally, Minerva Santos (2007) argues that while teacher peer evaluation in its current practice does not fulfill its original intent, it is useful in helping educators to better understand the practice of teaching.

The occurrence of these observations is somewhat ironic since colleges and universities are presumed to be the seat of higher education and learning. Academicians cannot seem to agree on what is an appropriate way of assessing the contribution of faculty members. Some institutions rely primarily on the evaluations from students. Others have used a combination of student evaluations and peer evaluations, but often based on different evaluation criteria. Others have relied on peer evaluations alone. Some institutions have used a summative approach based on the contribution of multiple peer evaluations, and sometimes self-evaluation; others have used a formative approach that attempts to look at the overall work of the faculty member. Whatever the approach, in many cases one of the following problems appear to exist:

a. The evaluation instrument is flawed or inadequate.
b. The evaluation process is flawed.
c. Both the evaluation instrument and procedure are flawed.

Because of these problems, the end result is often inaccurate and sometimes lead to unreasonable evaluations of the contribution of faculty members.

This paper examines the problem of teacher peer evaluation in colleges and universities. It commences by examining existing literature on the subject. It then highlights problems with existing peer review models by examining what various institutions have done. It then proposes a comprehensive teacher performance evaluation model that may be applied to any similar institution. The paper proceeds with five additional sections: section 2 conducts a brief review of existing literature on the subject matter; section 3 describes two cases related to the teacher peer evaluation models that have been implemented in two institutions; section 4 identifies problems with the two models; sections 5 proposes a teacher performance evaluation model (including a framework and an evaluation instrument) that can be adopted and applied to any institution; finally, section 6 summarizes and provides some concluding remarks.

2. LITERATURE REVIEW

The literature on teacher peer evaluation describes two forms of evaluation — summative evaluation and formative evaluation. Formative evaluation is a process that examines the performance effort of a team member, and provides feedback for the purpose of improving that performance. According to Suzan Allen Nan (2003), there are four types of formative evaluation — planning evaluation, implementation evaluation, monitoring evaluation and progress evaluation. Summative evaluation is a process that assesses the worth of a team member's professional contribution to an organization, with a view to using the information provided to inform decisions about promotion, increased compensation, or continued employment of the individual. Both formative evaluation and summative evaluation are important exercises, but they serve different purposes.

The Center for Effective Teaching and Learning (CETAL) at Texas University (2008) provides a useful comparison of formative evaluation with summative evaluation. Six points of divergence are identified: Firstly, formative evaluation is prospective (i.e. forward-looking); summative evaluation is retrospective. Secondly, formative evaluation examines strengths and weaknesses with a view to improving in the future; summative evaluation documents past performance and achievements. Thirdly, formative evaluation works towards developing habits; summative evaluation documents habits. Fourthly, formative evaluation shapes the direction of professional development; summative evaluation shows the results of such professional performance. Fifthly, formative evaluation provides feedback; summative evaluation looks at evidence. Finally, formative evaluation provides an opportunity to reflect on the meaning and significance of past achievements; summative evaluation provides evidence of formative evaluation and development.

2.1 General Considerations

In what has turned out to be to some extent, a springboard for discussions about improving the effectiveness of the peer review process, Larry Keig and Michael D. Waggoner teamed up to write an interesting paper (1994). In this paper, the authors argue that contrary to common presumption, summative evaluations seldom provide adequate information for improvement of faculty teaching. They suggest that formative evaluations may be more useful, but contend that both strategies should be employed. The article may be summarized in the following 8 submissions: First, Video-taping is the most effective form of classroom observation; other sources of information should include course materials, the instructor's assessment of students, and the teaching portfolios of instructors. Second, course materials are useful in assessing quality of preparation of the faculty member, which is often related to classroom performance. Third, faculty members who conduct peer evaluation should be appropriately trained; this is critical for the credibility of the exercise. Fourth, teacher participation is critical for the success of the peer evaluation program. Fifth, formative evaluations and summative evaluations should be treated as separate exercises. Sixth, formative evaluation should be nonjudgmental and informative, and should involve the observation of peers, administration, and students. Seventh, evaluation should be guided by experts in the particular discipline observed. Finally, the institution should provide incentives and rewards to encourage faculty members to participate in peer evaluation.

Matthew O. Richardson (2000) observes that most professors experience discomfort about a colleague being in their class for the purpose of evaluation. He also concurs with Keig and Waggoner that formative evaluations are more useful for improving instruction than summative evaluation. He recounts the experience of observing the class of a colleague that he considered senior to him and more accomplished. Initially, he felt uncomfortable at the thought of evaluating this accomplished colleague. As he shifted his focus from critique to learning how better to improve his own teaching, he became more comfortable, and the evaluation exercise became more meaningful. He argues that the intent of an evaluation should be to glean from the teacher, information that can benefit the observer. With this change of focus, from evaluation to observation, he proposes that the exercise allows teachers to glean from a wide variety of sources in an effort to improve their craft, fosters a sense of career-long learning, demonstrates to students that learning is an essential part of what professors do, and promotes a forum to talk about good teaching practices.

Rick Rudd, Tracey Hoover, Noelle Connor, and Matt Baker (2001) conducted a case study of the system of peer evaluations at the College of Agricultural and Life Sciences at the University of Florida. Among the findings of the study were the following: Firstly, there was inconsistency in how the evaluation was perceived by the faculty. Some chairs felt that the main purpose of the evaluation was to achieve and maintain excellence in teaching. Most chairs viewed it as an instrument used for decisions on promotion, tenure and awards. Most of the faculty without chairing responsibilities concurred with the latter view, but also viewed the evaluation as a means to enhance the learning experience of the students. Secondly, the faculty members also felt that although improving teaching was not the primary focus, it benefited from the exercise. Thirdly, there were concerns about the veracity of the information provided on evaluation reports, since the report was often skewed by the perception of its primary purpose. Finally, there were disparities and inconsistencies on the criteria used by different departments to evaluate teaching effectiveness.

While the benefits of peer evaluation are real, observe that there was some confusion among faculty members at the above-mentioned institution due the fact that a clear distinction between the two types of peer evaluations was not made. It appears that both formative and summative evaluation techniques were merged without careful thought and planning.

Another problem that often surfaces is the occurrence of unfair evaluations. Cyntia B. Schmeiser (1995) recognizes that there has been unethical conduct in assessments in areas such as use of test results, retention/promotion decisions, and accountability decisions. She notes that these problems are not new but date back to as early as the 1970s. Some of the reasons for unethical conduct in peer evaluations include failure to embrace diversity (Capella, 2005, p. 18), differing value systems among peers (pp. 72 - 73), groupthink and group-shift (pp. 217 -220), filtering, selective perception, and information overload (pp. 277 – 278).

Nancy Van Note Chism (1999) provides some very useful insights on peer evaluation. Regarding summative evaluation, she provides the following guidelines (pp. 33 – 34): Firstly, reviewers should make sure they are can be objective judges. If there are conflicts of interest, personal clashes with the evaluated individual, or any other issues that might prevent a fair evaluation, the prospective reviewer should excuse himself/herself from the evaluation team. Secondly, the conditions for a fair evaluation must be set. This means the absence of mistrust, and the presence of all needed information. Thirdly, reviews must be based on clearly established standards, and not personal preferences. Fourthly, conclusions must be based on facts and evidences, not opinions. Finally, the evaluation should focus on the overall performance of the teacher over the evaluation period, rather than given instances. Isolated findings should not determine the conclusions, but general patterns.

Chism draws on the work of Cohen and McKenzie (1980) in stating areas that should be evaluated: course goals and content, instructional method, evaluation of student' work, and interpretation of student ratings. She also cites French-Lazovick (1981, pp. 73 – 89) to recommend the following additional areas: quality of teaching materials, intellectual tasks set for students and the corresponding performance, knowledge of the subject area, assumption of the teaching responsibilities of the department and the institution, and commitment to excellence.

Chism proposes forms for peer review of course materials, covering broad areas including the course syllabus, course guidelines, formative evaluation instruments used for feedback from students, course-pack and textbook content, course bibliography, transparencies or presentation slides, handouts, multimedia course materials, tests, assignments, and assessment of students work (pp. 47 – 74). She also proposes forms for classroom observation, covering criteria including content knowledge, presentation skills, rapport with students, clarity, organization, instructional strategies, and impact on student learning (pp. 87 – 94).

The final item to be noted in this subsection is the work of Pat Hutchings (1996). In her scholarly work on the subject, she recommends that classroom observations should be done in groups, with each member visiting the classroom of the other members of the group (p.23). She also recommends the implementation of a mentoring program among faculty members as a means of promoting outstanding teaching (chap. 3). Finally, she argues that there should be a strong emphasis on the learning experience of the students (chap. 4).

2.2 University of Texas at Austin

The Center for Teaching Effectiveness (CTE) at University of Texas at Austin (n.d.) provides a comprehensive overview of the peer observation process. The document recommends that the peer observation exercise should be both formative and summative. It highlights the following areas for evaluation: mastery of course content, selection of course content, course organization, appropriateness of course objectives, appropriateness of instructional materials (i.e. readings, media), appropriateness of evaluative devices (i.e., exams, written assignments), appropriateness of methodology used

to teach specific content areas, commitment to teaching and concern for student learning, student achievement based on performance on exams and projects, and support of departmental instructional efforts.

The CTE document goes on to emphasize that in order for peer evaluation to be effective, training of faculty members in observation and analytical skills is necessary. It also underscores that care should be taken in developing the evaluation instrument. Three types of instrument are identified — checklist, rating scale, or written analysis. Two possible drawbacks identified with written analysis are that the analysis could be skewed and/or subjective. The CTE recommends that the following alternatives for formative peer observation:

- **Master Faculty:** Construct pairs of faculty members where each pair consists of a senior faculty member and a junior faculty member.
- **Mentor-Mentee Relationship:** Similar to the master faculty approach. The mentor provides friendship, emotional and professional guidance for the mentee.
- **Peer Development Triad:** Place faculty members in teams of three individuals.
- **Small Group:** Place faculty members in groups of four or more.
- **Graduate Student Feedback:** Have graduate students conduct peer observation of the professor's class.

As for the summative peer observation, the CTE suggests the following areas of observation: classroom performance observation, instructional materials review, advising activity review, participation on graduate committees and graduate teaching, special recognition for teaching, and overall evaluation.

Finally, the CTE document also provides guidelines and recommendations for formative evaluation as well as summative evaluation. Among these guidelines are the following: First, the peer observation system determined by a department must be well-understood by the participants. Second, observation/evaluation instrument must be carefully constructed. Third, all observation must be representative of overall teacher performance in the classroom. Fourth, each individual faculty observer should meet face-to-face with the individual instructor being observed. Fifth, the observer and the observed must be aware that observations do not take place in isolation; therefore, appropriate evidence must be gathered. Sixth, peer observation should be one aspect of summative evaluation. Peer observation alone should not dominate summative evaluation, since the former is often subjective and prone to inaccuracies. Seventh, precautions should be taken to minimize peer observation errors. Possible causes of errors include differing views among peers, inconsistencies in observation approaches among peers, reputation of the observed instructor, training of the observer, experience of the observer, the observation that colleagues tend to be better at assessing research than teaching service. Finally, peer observers/evaluators must be trained.

2.3 University of Adelaide

The Evaluation and Assessment Service of the Center for Learning and Professional Development (CLPD) at University of Adelaide (2008) outlines a summative approach to peer evaluation. The instrument used includes the following areas: knowledge of the teacher, scholarship in teaching, quality of teaching administration, awareness of formal regulatory matters, assessment of student learning, contribution to curriculum development and evaluation, other comments, and overall evaluation. The instrument used is fairly simple but comprehensive, and can therefore be easily expanded to be formative as well. Observers are invited to provide both qualitative and quantitative responses to questions in these areas.

Additionally, the university provides a set of forms that students complete. They come under the category of Student Evaluation of Learning and Teaching (SELT). There are three types of SELT forms: course evaluation form, the teacher evaluation form, and the program evaluation form. These forms are designed based on the Lickert scale. At the end of the year, the inputs from students are aggregated and this is used to provide useful information of the perception of the students on the courses, the performances of teachers, and the academic programs.

2.4 Indiana University and University of Iowa

The Campus Instructional Consulting department at Indiana University provides some useful guidelines and caution about peer evaluation. The document observes that peers are better positioned to judge the relevance of a teacher's course content, as opposed to students or teaching experts. It concurs with Chism that before any peer evaluation takes place, the evaluation criteria must be clearly defined and communicated. It also cautions that peer evaluation, when not conducted properly, could be the source of distrust among faculty members. It recommends that peer evaluation should be used only for discussion and improvement, and that for improved results, the observations should be done in pairs, or triads, with each member in a group observing each other member in the group.

The University of Iowa (2007) hosts a comprehensive system of faculty review, which is summarized here: Departments are given flexibility in planning and managing their respective systems of peer review, provided that they follow broad guidelines provided in the University's Operations Manual as well as the Faculty Handbook. Each faculty maintains a professional dossier of his/her performance. Evaluation of teaching effectiveness is facilitated by standard evaluations based on forms developed by the department. The process consists of student evaluations, classroom observations, and peer evaluations. The teacher has the right to respond, and this response forms part of the review file. In each department the peer review process is coordinated by the departmental executive officer (DEO) — a title used instead of the traditional department chair.

2.5 North Carolina University and Purdue University

The North Carolina State University (2006) recommends formative evaluation as well as summative evaluation. The former is used by each teacher to obtain feedback from students and peers in order to improve his/her teaching; the latter is used by the administration for make decisions relating to reappointment, promotion, tenure, teaching awards, and merit increase. The university has a student evaluation instrument that includes a set of core questions (determined by a committee and revised every three years), and other questions that the department or instructor may add. There are stringent guidelines to preserve the anonymity of students. Each department is responsible for constructing its peer evaluation instrument. However, the institution requires that each instrument covers some core issues such as resource materials used in the course, strategies used to promote learning, preparation and organization of the teacher, interaction with students, and observations specific to the discipline. Additionally, there are stringent guidelines for preserving the integrity of the peer evaluation process.

The School of Veterinary Medicine at Purdue University (n.d.) employs a rather interesting approach to peer evaluation. Peer evaluation teams of 4 or 5 individuals are set up. Each faculty member's teaching is assessed by a peer evaluation team (PET). The evaluation instrument includes criteria such as mastery of content, organization of the course, meeting of course objectives, appropriateness of resource materials, appropriateness of student evaluation methods, and commitment to teaching. Additionally, the university has stringent guidelines for preserving the integrity of the evaluation process. The university recommends three sets of meetings. The first meeting is to gather course materials. Secondly, the team members visit each other's classrooms. Thirdly, the entire team meets to discuss their observations and to arrive at consensus on each team member; a peer evaluation form is then filled out for each team member.

2.6 University of Saskatchewan and Cornell University

The University of Saskatchewan (2004) also joins overwhelming endorsement of both formative evaluation and summative evaluation. It also subscribes to the popular practice of giving respective schools and departments the latitude to develop their own evaluation instruments. Participation in peer evaluation is considered part of the faculty member's service to the institution. The university makes a distinction between peer consultation and peer evaluation. The former obtains feedback from peers, and is part of the formative evaluation exercise. The latter is used as part of the summative evaluation exercise. The university

recommends a minimum set of evaluation criteria including classroom performance, quality of examinations, course materials, syllabi, assignments, and other resource materials. It also recommends that peer evaluators should be of a higher rank than the evaluated individual.

In a similar but more elaborative manner, Cornell University (2007) places great emphasis on having a credible system of peer evaluation. The university's Teaching Evaluation Handbook provides useful guidelines. Among the more salient guidelines are the following:

a. Formative evaluation and summative evaluation should be kept separate since they serve different purposes. Also, keeping them separate will avoid conflict of interest between serving the needs of the teacher, and the needs of the administration (chap. 2).
b. Student data should be reported in summarized form; faculty members should not have access to the individual student evaluations (chap.3).
c. The effectiveness of the peer review depends on the effectiveness of the criteria by which the candidates are judged.
d. To be effective, the evaluation process should be neutral, open, unthreatening, and structured.
e. To avoid ambiguity, arbitrariness and subjectivity, clear questions based on established evaluation criteria must be developed.
f. Peer evaluation must be conducted by a team of individuals. Each evaluation must be conducted independently without consultation. The results should then be combined.
g. Individuals who conduct peer evaluations must be trained in this exercise. Unqualified individuals should not be given this responsibility.
h. Classroom observations should be conducted by individuals with no bias. Several observations should be conducted over time (chap. 4).
i. The evaluation must be grounded in facts and substance, not opinions and stylistics.
j. Regarding classroom observation, specific criteria are suggested for the following broad areas: class structure and goals, teaching behavior, rapport with students, subject natter and instruction. With respect to evaluation of course materials, criteria are specified for broad areas including course organization, course content, evaluation of student learning, course objectives, instructional methodology, and assignments. Regarding student evaluation, criteria are specified for the broad areas of analytic approach, instruction-group interaction, instruction-individual student interaction, and dynamism/enthusiasm.

3. TWO CASES OF PEER REVIEW MODELS

From the foregoing discussion, the following observations can be made: Firstly, there is no shortage of advice on the importance of teacher peer evaluation, and how it should be conducted. Secondly (and despite the previous observation), there is much inconsistency among institutions of higher learning as to how peer evaluation is actually conducted. Some institutions rely on summative evaluation alone; others bundle formative evaluation and summative evaluation into one; few treat formative evaluation and summative evaluation as separate issues. Some institutions develop and maintain an evaluation instrument based on predetermined criteria, and require each department to use it; others recommend that each department develops its own instrument in reference to some basic guidelines; others provide very vague guidelines and allow departments to conduct the evaluation however they choose. Some institutions conduct their summative evaluations based on student evaluations alone; others combine student evaluations with peer evaluations.

At this point, two peer evaluation cases will be presented and then analyzed. These cases were chosen because they provide several opportunities for discussion and for drawing useful lessons. In the interest of confidentially, fictitious institution names, departments, and individual names will be used.

3.1 Brainstorm Institute

Bruce Jones was employed by Brainstorm Institute (BI) to chair its Department of Environmental Science. Upon accepting employment at BI, Bruce noticed that there were two huge issues that needed urgent attention. Firstly, the environmental science (ES) curriculum was uncompetitive and therefore needed immediate revision. Secondly, the college relied on a student evaluation form as the only basis for assessment of the efforts of faculty members. The students were never trained on how to conduct such evaluations. This resulted in a high occurrence of evaluations that seemed inconsistent with the information that faculty members reported and/or expected.

Bruce assessed that these were huge challenges that needed focused attention. He decided to address the curricular issues first. It took him two academic years to plan and fully implement revised curricula in ES. Remedial courses were removed from the program, and replaced by additional college-level courses; also, more rigor was added to several of the preexisting courses. During this period of time, he and his departmental colleagues suffered very negative student evaluations from dissatisfied students who had become accustomed to the pre-existing status quo. They complained to the administration that the courses were too difficult. Several of those students dropped out and never returned to the institution. Bruce was forced to defend what he was doing at a meeting with all the department chairs and the Vice President for Academic Administration (VPAA). He explained to the administration that the credibility of the institution was at stake, demonstrated that he had a comprehensive marketing plan that would result in increased enrollment in the ES program over the medium term, and explained that this could not work with a mediocre program. The administration bought on to Bruce's plan with a high level of caution.

Bruce then turned his attention to the evaluation system employed by BI. He met with his team and explained the urgency of the situation to the members. After several brainstorming sessions and some research, they came up with a new peer evaluation system for their department. They developed 5 evaluation forms: Form 1 was the peer evaluation form; this allowed a faculty member to evaluate the work of his/her colleague based on clearly defined criteria in broad areas of professional preparation, classroom performance, team performance, and general observation. Form 2 was the self evaluation form; this allowed each faculty member to evaluate his/her own work based on the same criteria stated on form 1. Form 3 was for the evaluation of the department chair by members of the department; there were specific criteria in areas of leadership, competence, representation of the department, and general observation. Form 4 was the chair's self evaluation form, which allowed the chair to evaluate himself/herself based on the same criteria stated on form 3. Form 5 was the student evaluation form; this allowed students to evaluate each course based on clearly defined criteria in areas of course organization, teacher's preparation, resource materials, and learning experience. They then outlined and documented a procedure for conducting peer evaluations, as summarized below:

a. In each course, students were instructed on how to conduct a reasonable evaluation of a course. In particular, they were instructed to focus on the course, on the effort of the teacher, and on their personal involvement in the course.
b. Each faculty member was trained on how to conduct an objective peer evaluation and self-evaluation. In particular, they were trained to evaluate based on verifiable evidences and with respect to the established criteria, and not based on personal opinions, perceptions, or preferences.
c. Each faculty member's overall performance evaluation consisted of the aggregation of his/her peer evaluations and his/her own self-evaluation (on the same criteria), as well as the aggregation of student evaluations on each course taught by that faculty member over the evaluation period.
d. The department chairperson's overall performance evaluation consisted of the aggregation of peer evaluations by team members, and the chairperson's own self-evaluation (on the same criteria); thus, the chairperson had two overall performance evaluations — one as a faculty member, and the other as chairperson.

To further preserve the integrity of the system, Bruce and his team configured spreadsheets that would allow for independent entry of individual evaluations (from peers as well as students), and the generation of the aggregate reports. Each faculty member would be given a copy of the aggregate report relating to his/her performance; a copy would also be placed on the faculty member's file.

This system of evaluation worked quite well for Bruce's department. Two years after its implementation, Tom Glavin, a faculty member in the department was awarded the teacher of the year. Moreover, other departments learned about the system, and expressed an interest in adopting it — an outcome that Bruce and his team were happy to witness.

3.2 Lehman Polytechnic State University

Scott Chisolm entered the employ of Lehman Polytechnic State University (LPSU) with several years of experience and an impressive track record as an outstanding environmentalist and teacher of environmental science. Upon accepting employment at LPSU, Scott noticed that there were two huge issues that needed urgent attention. The environmental science (ES) curriculum needed revision in order to make it more competitive and relevant to the needs of students. Secondly, the university relied on a student evaluation form, coupled with a loose system of peer evaluation without clearly defined criteria, as the basis for assessment of the efforts of faculty members. Without clear guidelines, peer evaluation teams were at liberty to conduct subjective qualitative evaluations of their peers, based on their own perceptions — such evaluations could be overly positive or overly negative of a colleague. Additionally, there was no evidence that students were appropriately trained on how to conduct course evaluations, or that faculty members were trained on how to conduct peer evaluations.

Scott decided to address the curricular issues first, and with time raise attention to the peer evaluation problem. His preliminary attempt at introducing changes to the ES curriculum was successful. However, his attempts to make more significant changes were met with opposition. The other members of the department became overly defensive because they thought Scott was debunking their previous efforts. To complicate matters, they had several things in common such as ethnicity, and shared values. Scott was of a different ethnicity from the other departmental members, and embraced some different values from them. For instance, Scott had a strong team-work orientation, while his team-members tended to be more individualistic; he had a perfectionist approach to work, while his team members were often less extreme and sometimes seemed willing to settle for above average.

Scott was very thorough in his teaching. He challenged his students to strive for excellence, provided them the resources needed to succeed, and insisted that they make the effort. He often arranged extra tutorials, and would stay late evenings for those who needed extra help. The reaction of students to Scott's teaching was mixed. Some of the senior students did not like the fact that they were being challenged to lift the standard of their performance, and they complained profusely; others were excited and felt that this was a positive change for the department. The first year's peer evaluation reflected this mixed excitement about the changes in the department, and was in general, fairly positive.

The second year was much busier for Scott. Due to limited departmental resources, he carried teaching overloads in both semesters, and served on three standing committees at the university. Halfway through the second semester, he again brought up the matter of curriculum review, and expressed concerns that the department was not addressing the matter with sufficient urgency; he argued that this was not in the best interest of the students. This was met with intense opposition from some faculty members of the department, who apparently felt that their previous efforts were being debunked by Scott. Scott tried to explain that his focus was not that of disparaging the efforts of anyone, but rather to use prior work as a springboard for further refinement in the best interest of the department and its students. Still, negative sentiments were high. After a heated debate, he was informed by Karen Campollo, the ES chair, that the other members did not see any need to improve the ES curriculum. Shortly afterwards, he received e-mails from Karen (sent to all department members) that implicitly accused him of trying to undermine the stability of the department for personal gain. Shocked and concerned, he went to the Dean of Science, Rushmore Frantz, explained the situation to him, showed him the e-mails, and expressed concerns about possible reprisal on his upcoming peer evaluation, due to his effort to improve the quality of the department's curriculum and services to its students. Frantz offered some advice, and the two agreed to monitor the situation.

Then came the peer evaluation. Scott's peer evaluation committee (PEC) consisted of Karen Campollo (as ES chair), Cynthia Carty (a department member and close friend of Karen), and Erron Lazarus (an observer from another department). The understanding was that Karen would be the PEC chair. Statistically, the student evaluations were an improvement over the previous year. He had a 100% pass-rate in 80% of the courses taught, and an average of 90% in the other courses; he also received an increased level of net positive comments on the student evaluations; more significantly, he had earned the respect of the students in the department. Extraordinarily, the peer evaluation was significantly worse than the previous year. The peer evaluation made two commendations: it commended him for being able to add humor to his classes, and being able to adjust to the needs of the students. It then went on to make charges including unclear presentations, fast pace of teaching, unrealistic student expectations, focus on the weaknesses of his students, high drop-out rates, and not meeting students at their level. The peer evaluation also criticized his use of statistics to convey student evaluation and performance as being repetitive, made claims of recurring concerns on issues that were never mentioned before, and charged him of not providing course documentations that were verifiably provided. Finally, on the basis of these allegations, the report recommended that Scott should not be promoted, despite the fact this was understood to be a regular peer evaluation and not an assessment on Scott's fitness for promotion. (Scott had the option of applying for promotion, but had decided to defer that option at the time.)

Scott prepared a detailed response to the PEC report, and on each allegation, showed its inconsistency or inaccuracy. He also showed where the PEC failed to report verifiable facts that would have portrayed his work in a more positive manner, reported personal opinions as facts, and drew conclusions based on those personal opinions. Scott also observed that despite his availability, the PEC made no noticeable attempt to obtain clarification on the pertinent issues either before or after the writing of the report, but was apparently enthusiastic in its misrepresentation of those very issues. He then met with Rushmore to discuss his options. Rushmore advised him to make some adjustments to his response, and to seek a meeting with the PEC. Scott met with Erron Lazarus, to inquire what might have prompted the extraordinarily negative PEC report. Erron mentioned that he was the author of the report, went on to repeat the negative accusations of the PEC report, and added a few new ones of disingenuousness, student bashing, and possible fudging of student grades. Scott inquired of the Erron if he was aware that the allegations were inconsistent with verifiable facts about his work. Erron responded to the effect that the report was based on what was discussed among the PEC members. He also mentioned with a rather gloating attitude, that the other two PEC members told him that the report would be more effective if he wrote it, so he did. Further pressed by Scott, Erron eventually admitted that he did not understand some of the issues related to Scott's peer evaluation. After a discussion that lasted over 3 hours without any agreement, Scott thanked Erron for his time, and pointed out that there were fundamental differences between them on the issues.

At a second meeting between the two, Erron surprised Scott by hurling some personal attacks and insults at him, making judgments about his personality and character. At that point, Scott determined that it was time to permanently end discussions between them. He indicated to Erron that on the basis of the foregone exchanges, no further discussion between them was necessary.

Scott reported back to Rushmore on his meetings with Erron, revised and submitted his response of objection to the PEC report. In his closing remarks, he pointed out that by reason of its inaccuracies, inconsistencies, misrepresentations, and omissions, the PEC report lacked credibility. He subsequently brought the issue to the attention of Lambert Farnsworth, the newly appointed ES chair. After further probing and contemplation of the issue, Scott informed Lambert and Rushmore that he had come to the conclusion that the PEC report was scandalous and unethical, and should be treated as such. However, in the interest of departmental unity and student morale, he chose not to press the matter any further.

4. OBSERVED PROBLEMS

The two cases in the previous section provides us with compelling reasons educational institutions should exercise unremitting care in ensuring and preserving integrity (Paine, 1994) in their systems of peer evaluation. The problem of inconsistency in standards for peer evaluation has already been noted. Let us now examine the two previously-mentioned cases and see how this problem can mushroom into several others.

4.1 Analysis of the BI Case

In the case of euphemistic institution called BI, 3 problems can be identified:

a. The institution had no system of formative evaluation in place. This meant that faculty members had no system of objectively evaluating their work with a view to improving their performance. This can often lead to stagnation, and a false sense of achievement.
b. There was total reliance on student evaluations as the basis for summative evaluation. The problem with this approach is that students are not qualified to make judgment of course quality, particularly in areas of content, organization, and adequacy of resources materials. If they were, they would not be pursuing such courses. Their observation is important, but should not be the sole source of evaluation.
c. Lack of training of students on how to conduct course evaluations might have resulted in the high occurrence of skewed course evaluations. The assumption is often made that students know how to conduct such evaluations. That assumption is not supported by any evidence. In most cases, students actually need guidance in this area.

Bruce and his team were able to address these problems and achieve the success they did. Despite the fact that they did not clearly differentiate between formative and summative evaluation, but rather merged the two, they achieved success because they had other constraints in place to preserve the integrity of their system. These constraints included the following:

- Evaluation criteria that were clearly defined and communicated to the stakeholders (faculty members and students)
- Clearly defined procedures and guidelines for the stakeholders
- The design and implementation of a set of evaluation instruments that included student evaluations, peer evaluations, and self evaluations that were based on the established criteria, and constrained by the procedures and guidelines

4.2 Analysis of the LPSU Case

The LPSU case is even more troubling than that of BI. A number of problems can be identified:

a. No clear distinction was made between formative evaluation and summative evaluation.
b. There was an absence of clearly defined criteria for evaluating teaching effectiveness. This opened the door for the kind of unfair evaluation that Scott became the victim of.
c. As in the case of BI, lack of training of students on how to conduct course evaluations might have resulted in a high occurrence of skewed course evaluations.
d. Faculty members were not trained on how to conduct peer evaluations. The assumption that they knew how to do so turned out to be incorrect in this particular instance.
e. The conduct of the PEC was at best questionable on a number of issues. Failing to clarify the issues, and then having Erron write a report that should have been written by Karen, raises a number of flags. Firstly, according to Erron's own admission, he wrote the report because the other two PEC members felt it would be more effective if he did. Erron was not qualified to write such a report, and as he subsequently admitted to Scott, he did not understand some of the critical issues since he was not an ES professional. This suggests that there was some collusion on the matter. Secondly, consistent with standard requirement for suitability (Chism, 1999, p.33), if Karen knew that she that for whatever reason she was unable to assume her designated role on the PEC (which apparently she did), she should have excused herself from the exercise.
f. In the closing arguments of his response, Scott highlighted the inaccuracies, inconsistencies, misrepresentations, and omissions of the PEC report as the basis for its lack of credibility; he subsequently concluded that the report was unethical. As mentioned in section 2, the conclusions of a peer review should be grounded in verifiable facts, not individual perceptions or opinions. With this benchmark, it appears that Scott was unfairly evaluated based on criteria not clearly articulated. This suggests that he might have been penalized due to his difference. Such action violates the principle of diversity (Capella, 2005, p.18) and is discriminatory (Velasquez, 2006, chap. 7).
g. The treatment meted out to Scott appears to violate ethical theories of utilitarianism, rights, justice, and virtue. A full discussion of ethics is beyond the scope of this paper, but a brief elaboration is in order: The utilitarian principle is about maximizing utilities for the community served (Bentham, 1781); it was violated here because in wrongfully penalizing Scott, the PEC threatened the continued willing service of a colleague who appeared to be very committed to his profession, and the department that he was serving. The principle of rights holds that all human beings have rights from which they should not be denied (Sullivan, 1989; Velasquez, 2006, pp. 71 – 84); Scott's right to a fair evaluation was violated. Justice relates to being right, just and fair (BrainyQuote.com, n.d.; Rawls, 1999; Velasquez, 2006, pp. 89 – 99); the treatment that Scott received was neither right, just, nor fair. Virtue relates to exemplary human conduct based on exemplary character (Stanford Encyclopedia of Philosophy, 2007); the conduct of the PEC fell short of this benchmark.

In section 2.1, it was mentioned that unethical conduct in peer evaluations could result due to factors such as failure to embrace diversity, differing value systems among peers, groupthink and group-shift, filtering, selective perception, and information overload. It appears that almost all of these factors were at play in this case. Additionally, the absence of clearly defined evaluation criteria, procedures, and guidelines, contributed to a rather undesirable situation which resulted in the unfortunate evaluation of Scott's work.

5. PROPOSED TEACHER PERFORMANCE EVALUATION MODEL

The foregoing discussions emphasize the significance of *teacher performance evaluation*, and importance of preserving the integrity of the exercise (see Paine, 1994). Notice the deliberate shift in emphasis from *peer evaluation* to *performance evaluation*. The latter term is much broader (and more preferred by this paper), including peer evaluation, as well as other aspects such as student evaluation and self evaluation (the acronym TPE is retained but with the understanding that the "P" here stands for "performance"). Failure to ensure and preserve the integrity of the TPE system is worse than not having one in the first place. This is so for a

number of reasons: Firstly, flawed teacher performance evaluations are unethical and therefore have no place in an institution with the mission of providing education to the public. They set the wrong example to students as well as teachers. Secondly, flawed teacher performance evaluations can lead to misrepresentations of the professional effort of teachers. Overly positive evaluations could lead to lethargy and failure on the part of the recipient to improve teaching quality; overly negative evaluations could lead to de-motivation and possible turnover, thus affecting the stability of the workforce. Thirdly, flawed teacher performance evaluations can erode trust and respect among members of a team. Fourthly, as a consequence of its possible effect on trust, motivation, and stability of the workforce, flawed teacher performance evaluations can lead to a loss in productivity.

Against this background, this paper proposes a generic performance evaluation model that may be applied to any college or university. Notice the deliberate shift in emphasis from *peer evaluation* to *performance evaluation*. The latter term is much broader (and more preferred by this paper), including peer evaluation, as well as other aspects such as student evaluation and self evaluation (the acronym TPE may be retained but with the understanding that the "P" here stands for "performance"). Also note that the model is inclusive of teaching and non-teaching members of the delivery system for the institution. The model consists of a framework and a set of evaluation instruments.

5.1 Framework

Based on the various sources discussed in section 2 (for instance, see section 2.1), and the cases discussed in sections 3 and 4, this paper proposes a framework for performance evaluation based on 12 fundamental principles. These principles are described in figure 1. They are the evaluation scope, distinction between formative evaluation and summative evaluation, selection of evaluation criteria, training of faculty, training of students, factual evidences, focus on the big picture, transparency, student anonymity, team involvement, sound judgment, ethical integrity, and ownership.

In addition to the above-mentioned principles, it is recommended that the principle of *management by objective* (MBO) be employed. This principle prescribes a methodical approach to defining corporate and individual objectives, assigning targets, evaluating work, and rewarding effort at each management level in the organization (Capella, 2005, pp. 164 – 166). Based on this principle, the following guidelines are proposed:

a. At each management level of the institution, clearly defined objectives are agreed upon in accordance with the mission of the institution and the mission of that particular organizational unit. Specific targets are set over determined performance cycles for each employee. The term *performance cycle* is used here to mean the period over which the individual's effort will be evaluated.
b. At each management level of the institution, clearly defined criteria for evaluation of performance are defined and communicated.
c. The evaluation for the president (or principal leader of the institution) consists of a peer evaluation by each vice president (next in command), a self evaluation, and an evaluation by the board.
d. At every management level, the evaluation for an administrator (for example head of department) consists of an evaluation by his/her immediate supervisor, peer evaluations by his/her team members, and a self-evaluation.
e. Each teacher's evaluation consists of peer evaluations by members of a peer evaluation committee approved by his/her department, a self-evaluation, and student evaluations.
f. Each remaining non-teaching employee's evaluation consists of an evaluation by his/her immediate supervisor, peer evaluations by his/her team members if applicable, and a self-evaluation.

In these proposed guidelines, a distinction is made between a supervisor and a peer. This distinction is purely for convenience, since a supervisor could very well be a peer that is fulfilling this special role in the institution.

TABLE 1: Proposed Fundamental Principles of Teacher Performance Evaluation

Principle	Description
01. Evaluation Scope	TPE should include formative evaluation and summative evaluation.
02. Separation	It is recommended that formative evaluation and summative evaluation be treated as separate issues. However, if due to limited resources, the two are merged, unremitting care should be taken to ensure that the integrity of the exercise is not compromised.
03. Criteria	TPE should be contingent on clearly defined criteria. An evaluation instrument must be developed based on the defined criteria.
04. Faculty Training	Faculty members must be trained on how to conduct peer evaluations. Unqualified individuals should not participate in the evaluation exercise.
05. Student Training	Students must be trained on how to conduct course evaluations. This can be done as part of the student's orientation, or built into certain core course(s) that the student must take.
06. Evidence	TPE reports must be grounded in factual evidences, and never based on individual opinions or perceptions.
07. Big Picture	TPE reports must be based on overall performance over the evaluation period, and not on isolated incidents.
08. Transparency	The evaluation process must begin and end with a face-to-face meeting so that needed clarifications can be obtained, and trust can be reinforced.
09. Student Anonymity	Anonymity of the participating students must be preserved. Faculty members must not be given raw student data, but the summarized form of the data.
10. Team Involvement	TPE exercises should be conducted by a team of individuals. Each evaluation must be conducted individually without consultation with other team members. The results should then be combined with the candidate's self-evaluation, to yield an overall evaluation.
11. Sound Judgment	Peer evaluators must exhibit sound judgment and the ability to be neutral. A prospective evaluator should excuse himself/herself from the exercise, if he/she believes there are circumstances that would compromise sound and unbiased judgment of the candidate's effort.
12. Ethical Integrity	TPE must be conducted in an environment of trust and integrity, not distrust and speculation.
13. Ownership	Evaluation criteria and guidelines must be discussed and agreed upon by participating members, thus giving a sense of ownership to the members.

5.2 Evaluation Instruments

Again drawing from the literature covered in section 2 (see subsections 2.2 – 2.6), this paper further proposes a set of 5 evaluation instruments. The first is the Peer Evaluation of Teacher instrument and is based on criteria specified in table 2. Next is the Teacher Self Evaluation instrument, which is also based on the same criteria expressed in table2; the difference here is that the teacher conducts a self-evaluation of his/her work. The third instrument is the Peer Evaluation of Department Chair instrument, based on criteria specified in table 3. Next is the Department Chair Self Evaluation instrument, based on the same criteria as specified in table3; the difference here is that the department head would be required to conduct a self-evaluation of his/her work. The fifth instrument is the Course Evaluation instrument, based on criteria specified in table 4.

As table 2 reflects, the recommended criteria for peer evaluation of the teacher are placed in ten categories: professional preparation, classroom performance, course quality, observed learner experience, evaluation of student learning, interaction with students, scholarship, concern for teaching, team and community performance, and general observations. In the interest of integrity and fairness, it is recommended that these criteria be held for the peer evaluations as well as the teacher's self-evaluation; failure to do this diminishes the exercise. It is also recommended that except for the last three questions which are open-ended, the questions should be scaled via the Likert scale [Trochim 2006], with appropriate clarification for the range of possible values on each question. In the table, the suggested response to each scaled question is one of the following: *strongly agree* (SA), given a nominal value of 5; *agree* (A), with nominal value of 4; *neutral* (N), with nominal value of 3; *disagree* (D), with nominal value of 2; *strongly disagree* (SD), with a nominal value of 1. Finally, to obtain a reasonable evaluation of the teacher's performance, an aggregation of the peer evaluations and the teacher's self-evaluation should be taken.

As table 3 reflects, the recommended criteria for peer evaluation of the department chairperson are placed in two categories: leadership and competence, and general observations. Again, in the interest of integrity and fairness, it is recommended that these criteria be held for the peer evaluations of the department chair as well as the chair's self-evaluation. The last three questions are to be open-ended, and the others based on the Likert scale. A final read of the chairperson's performance should be obtained by aggregating the peer evaluations and the chairperson's self-evaluation.

Table 2: Evaluation criteria for teacher

		SA (5)	A (4)	N (3)	D (2)	SD (1)
A.	**Professional Preparation**					
01.	Prepares detailed course based on departmental standards					
02.	Makes course materials interesting and relevant to student					
03.	Carefully prepares lectures					
04.	Exhibits mastery of subject matter taught					
05.	Keeps abreast of current developments in related professional areas					
B.	**Classroom Performance**					
06.	Clearly states class objectives					
07.	Starts class on time					
08.	Communicates clearly and effectively					
09.	Exhibits enthusiasm for the subject matter					
10.	Imparts confidence to students					
11.	Uses time efficiently					
12.	Explains concepts clearly					
13.	Focuses on the needs of students					
14.	Consistently takes class attendance					
15.	Arranges class tutorials if and when required					
16.	Responds appropriately to questions from students					
17.	Organizes material to meet learning objectives of the class					
18.	Encourages critical thinking and analysis					
19.	Makes effective use of resources					
20.	Employs effective teaching practices to encourage learning					
21.	Employs appropriate illustrations to reinforce concepts					
22.	Demonstrates personal interest in students					
C.	**Observed Quality of Course(s) Taught**					
23.	Prepares detailed course syllabus for each course					
24.	Syllabus includes clearly and appropriately defined objectives					
25.	Content is appropriate for each course					
D.	**Observed Learner Experience**					
26.	Learners Seemed motivated and engaged					
27.	Learners appeared comfortable with the material presented					
28.	Learners appeared comfortable with the pace of the course					
E.	**Evaluation of Student Learning**					
29.	Assignments and tests are appropriate for the course(s) observed					
30.	Other used means of student assessment are appropriate for course					
31.	Instructor maintains verifiable record of student performance					
F.	**Interaction With Students**					
32.	Shows concern for students					
33.	Encourages participation from students					
34.	Available for students outside of regular class time					
E.	**Observed Quality of Course(s) Taught**					
29.	Prepares detailed course syllabus for each course					
30.	Syllabus includes clearly and appropriately defined course objectives					
31.	Content is appropriate for each course					
F.	**Observed Learner Experience**					
32.	Learners Seemed motivated and engaged					
33.	Learners appeared comfortable with the material presented					
34.	Learners appeared comfortable with the pace of the course					
G.	**Scholarship**					
35.	Demonstrates evidence of significant research involvement					
36.	Demonstrates evidence of creative original work					
37.	There is evidence of others using his/her work					
38.	Demonstrates evidence of being well read					
39.	Demonstrates evidence of current awareness in the related field(s)					
H.	**Concern for Teaching**					
40.	Seeks advice from others about teaching courses he/she teaches					
41.	Discusses teaching with colleagues					
42.	Exhibits care and interest about hi/her teaching					
I.	**Team and Community Performance**					
43.	Integrates and collaborates well with other department members					
44.	Treats with respect the views of other members of the department					
45.	Respects and upholds the mission of the department					
46.	Contributes to the achievement of departmental objectives					
47.	Participates in activities to improve/support the community					
J.	**General Observations**					
48.	Main strengths of the instructor					
49.	Main weaknesses of the instructor					
50.	Other comments					

Table 3: Evaluation Criteria for Department Chair

A. Leadership and Competence	**SA (5)**	**A (4)**	**N (3)**	**D (2)**	**SD (1)**
01. Involves department members in decision-making					
02. Keeps department members informed relevant matters					
03. Promotes and encourages departmental communication and unity					
04. Handles departmental matters well					
05. Is sensitive to faculty morale					
06. Confers with the department members regarding the departmental budget					
07. Stimulates faculty members to become involved in professional development					
08. Develops a satisfactory teaching schedule for department members					
09. Evaluates teachers and courses consistently and fairly					
10. Develops and implements an effective system of student advising					
11. Provides exemplary leadership for the department					
12. Implements timely evaluation of the department's program curriculum/curricula					
13. Demonstrates and encourages professional excellence					
14. Keeps abreast of developments within the field(s) represented					
15. Keeps abreast of developments in higher education					
16. Is accessible and helpful to members of the department					
17. Represents the department appropriately within the community of the institution					
18. Manages conflicts effectively					
19. Encourages differing views on critical issues					
20. Manages change satisfactorily					
B. General Observation					
21. Main strengths of the chair					
22. Main weaknesses of the chair					
23. Other comments					

Turning to course evaluations (table 4), the recommended criteria span five categories: the course, the teacher's performance, the teacher's interaction with students, the learning experience of the students, and general comments. It is further recommended that each course be evaluated based on three data sets, each based on the same criteria: the first is the student evaluations of the course; the second is a set of peer evaluations of the course; the third is a self evaluation of the course by the faculty member. These data sets should then be weighted and combined into a single report. It is imperative that these data sets be based on the same evaluation criteria; otherwise, the credibility of the report is significantly compromised. Again, the Likert scale is recommended except for questions in the general comments area.

Observe that these instruments can be easily automated in a software system. In a less sophisticated scenario, they can be implemented with a spreadsheet program (such as Microsoft Excel or Quattro Pro). It should also be noted that for each evaluation instrument, the institution has the option of selecting and clarifying the criteria that are pertinent to its mission and local situation. Moreover, depending on departmental preferences, some of these criteria may vary among the various academic departments of the institution.

Table 4: Course Evaluation Criteria					
A. The Course	**SA (5)**	**A (4)**	**N (3)**	**D (2)**	**SD (1)**
01. The objectives and requirements are clearly defined					
02. The syllabus clearly outlines the sequence of topics					
03. The assigned textbook(s) is/are useful					
04. The assignments and tests were challenging					
05. The resource materials and/or reading assignments were useful					
B. The Teacher					
06. Is typically on time for classes					
07. Is typically well prepared					
08. Stimulates my interest in the subject					
09. Demonstrates the ability to explain and clarify the material covered					
10. Provides different perspectives to the subject matter discussed					
11. Relates recent developments on important aspects of the course					
12. Demonstrates consistency and fairness on evaluating students' work					
13. Gives clear examples					
14. Is careful and precise in answering questions from students					
15. Systematically summarizes major sections of materials covered					
16. Provides adequate feedback to students					
17. Demonstrates enthusiasm for the subject matter					
18. Appears to enjoy teaching					
19. Exudes self confidence					
20. Exhibits a good sense of humor					
21. Is considered by me to be an excellent teacher					
C. Teacher Interaction with Students					
22. Encourages and is responsive to student participation					
23. Demonstrates genuine interest in students					
24. Invites criticisms of his own ideas					
25. Detects when students are confused or bored					
26. Recognizes students outside of the class					
27. Is accessible outside of normal class times					
28. Provides useful advice on matters other than the course					
29. Shows respect for students					
D. Learning Experience					
30. I enjoyed the course					
31. My interest in the subject has increased					
32. The course was challenging					
33. I would recommend this course to others					
E. Other Comments					
34. What did you like most about the course?					
35. What would you suggest to improve the quality of the course?					

6. SUMMARY AND CONCLUDING REMARKS

Here is a summary of what has been presented in this paper: Teacher performance evaluation in institutions of higher learning is of paramount importance. However, there seems to be a lack of standards as to how this exercise should be conducted. This lack of standards has led to proliferation of varying approaches in different institutions. This inconsistency often leads to inadequate evaluations and/or unethical evaluations, neither of which is beneficial to an educational institution.

While both formative evaluation and summative evaluation are important components of the faculty performance evaluation process, distinguishing between them is equally important in avoiding ambiguities, and preserving the integrity of the exercise. In order to elevate the performance evaluation exercise to an increased level of integrity and efficiency, this paper has proposed a twelve-principle framework, an MBO-based approach characterized by clearly defined objectives and evaluation criteria, and a set of evaluation instruments based on the defined criteria. This performance evaluation model can be applied to any institution. It is also expected to contribute positively to improved job satisfaction, trust and productivity among participants. This could translate into huge benefits to the host institution.

7. REFERENCES

Alicias, E. R. 2005. Toward an objective evaluation of teacher performance: the use of variance partitioning analysis, VPA. Education Policy Analysis Archives, 13, 30, 115.

Bentham, J. 1781. Introduction to the principles of morals and legislation. http://www.utilitarianism.com/jeremy-bentham/index.html, accessed January 2007.

BrainyQuote.com. n.d. Justice. http://www.brainyquote.com/words/ju/justice181970.html, accessed January 2008.

Braskamp, L. A. & Ory, J.C. 1994. Assessing faculty work. San Francisco: Jossey-Bass.

Bushman, J. 2006. Teachers as walk-through partners. Educational Leadership, 63, 6, 58-61.

Capella University. 2005. OM8004: Managing and organizing people (an edited work). Boston, Prentice Hall Custom Publishing.

Center for Effective Teaching and Learning (CETaL) at Texas University. 2008. Formative and summative evaluation. http://sunconference.utep.edu/CETaL/resources/portfolios/form-sum.htm, accessed March 2008.

Chism, N. V. N. 1999. Peer review of teaching: a sourcebook. Bolton, MA: Anker Publishing.

Cohen, P. & McKenzie, W. 1980. The role of colleagues in the evaluation of college teaching. Improving College and University Teaching, 28, 4, 147 – 154.

Cohen, J. 2003. Documenting teacher excellence within the peer review of teaching. Journalism and Mass Communication Education, 58, 2, 115-118.

Cornell University. 2007. Cornell University teaching evaluation handbook. http://www.clt.cornell.edu/resources/teh/intro.html, accessed February 2008.

French-Lazovick, G. 1981. Peer review: documentary evidence in the evaluation of teaching In J, Millman (Ed.), Handbook of teacher evaluation. Thousand Oaks, CA: Sage.

Hutchings, P. 1996. Making teaching community property: a menu for peer collaboration and peer review. Washington, DC: AAHE Teaching Initiative, American Association for Higher Education.

Indiana University. 2006. Peer evaluation and teaching. http://www.indiana.edu/~teaching/allabout/pubs/peer.shtml, accessed February 2008.

Keig, L. & Waggoner, M. D. 1994. Collaborative peer review: the role of faculty in improving college teaching. Washington DC: George Washington University. http://www.ntlf.com/html/lib/bib/94-2dig.htm, accessed January 2008.

Murray, C. E. & Grant, G. 1998, Summer. Teacher peer review: possibility or pipedream? Contemporary Education, 69, 4, 202 – 203.

Nan, S. A. 2003. Formative evaluation. Boulder: University of Colorado. http://www.beyondintractability.org/essay/formative_evaluation, accessed March 2008.

North Carolina State University. 2006. Evaluation of teaching. http://www.ncsu.edu/policies/employment/faculty/REG05.20.10.php, accessed February 2008.

Odden, A. 2004. Lessons learned about standard-based teacher evaluation systems. Peabody Journal of Education, 79, 4, 126-137.

Paine, L. S. 1994, March. Managing for organizational integrity. Harvard Business Review, 72, 2, 106 – 117.

Peterson, K. 2004. Research on school teacher evaluation. NASSP Bulletin, 88, 60-79.

Purdue University. n.d. Peer teaching evaluation. http://www.vet.purdue.edu/PeerTeachingEvaluation, accessed February 2008.

Rawls, J. 1999. A theory of justice (Revised edition). Oxford University Press.

Richardson, M. O. 2000, Summer. Peer observation: Learning from one another. National Education Association Thought & Action Journal, Summer, 9 - 20. http://www2.nea.org/he/heta00/images/s00p09.pdf, accessed February 2008.

Rudd, R., Hoover, T., Conner, N. and Baker, M. 2001. Peer evaluation of teaching in University of Florida's College of Agricultural and Life Sciences. Journal of Southern Agricultural Education Research, 51, 1, 189 – 200.

Schmeiser, C. B. 1995. Ethics in assessment. http://ericae.net/db/edo/ED391111.htm, accessed February 2008.

Stanford Encyclopedia of Philosophy. 2007, June 14. Plato ethics: an overview. http://plato.stanford.edu/entries/plato-ethics/, accessed January 2008.

Stanford Encyclopedia of Philosophy. 2007, July 17. Aristotle ethics. http://plato.stanford.edu/entries/aristotle-ethics/, accessed January 2008.

Stanford Encyclopedia of Philosophy. 2007, July 18. Virtue ethics. http://plato.stanford.edu/entries/ethics-virtue/, accessed January 2008.

Santos, M. 2007, Spring. Peer evaluations and the language classroom. Academic Exchange Quarterly, 11.1, 71, 6.

Sullivan, R. J. 1989. Immanuel Kant's moral theory. New York: Cambridge University Press, 1989.

Trochim, W. M. K. 2006. Likert scaling. http://www.socialresearchmethods.net/kb/scallik.htm, accessed March 2008.

University of Adelaide. 2008. Peer evaluation of teaching. http://www.adelaide.edu.au/clpd/evaluation/pet.html, accessed February 2008.

University of Iowa. 2007. Faculty review. http://www.clas.uiowa.edu/faculty/review/index.shtml, accessed February 2008.

University of Saskatchewan. 2004. Instructional development. http://www.usask.ca/university_council/idcc/reports/06-19-03.shtml, accessed February 2008.

University of Texas at Austin. n.d. Preparing for peer observation: a guidebook. http://www.utexas.edu/academic/cte/PeerObserve.html, accessed February 2008.

Velasquez, M.G. 2006. Business ethics: concepts and cases. Upper Saddle River, NJ: Prentice Hall.

Yon, M., Burnap, C. & Kohut, G. 2002. Evidence of effective teaching: perceptions of peer reviewers. College Teaching, 50, 3, 104-110.

Weimer, M. 1990. Improving college teaching: strategies for developing instructional effectiveness. San Francisco: Jossey-Bass.