Re-conceptualising approaches to the evaluation of teaching quality
Zenawi Zerihun, Willem van Os and Jos Beishuizen

Abstract
Higher education institutions in many countries have been introducing a system of evaluating the quality of teaching in response to pressure from governments and the public. Such a system has, however, been influenced by lack of acceptable definition of what constitutes effective teaching. Even when attempts are made at defining the construct, more emphasis is given to evaluating the effectiveness of teacher characteristics. Students’ course grades are also used as indicators of student learning, hence the effectiveness of the teaching. Although both teacher characteristics and course grades can be considered as indicators of the effectiveness of teaching, they are influenced by extraneous factors. Hence, other sources, such as the students’ perceptions of their learning should be used to complement the data from course grades and teacher characteristics. Moreover, it is equally important to put in place a system of checking the appropriateness of the evaluation system for the stated purpose, either based on the interrelationship of the teacher evaluation with other measures of teaching effectiveness or looking at the improvements teachers make following the feedback. As the conceptions of teaching and learning held by teachers and students influence their judgments of the effectiveness of teaching, it is important to provide training in such a way that both can have similar conceptions. Finally, it is also appropriate for the teaching evaluation system to have a clear purpose of conducting the evaluation, i.e. summative or formative evaluation. The implications of the framework to the quality improvement efforts in Ethiopian universities are discussed.

Introduction
The quest for quality teaching has been on the higher education agenda in recent years. In Ethiopia, responding to ever increasing demands for quality from students and other stakeholders, higher learning institutions have been attempting to introduce quality control as well as enhancement mechanisms; these focus either on the quality of the educational program, the quality of performance of the provider, or the quality of the students’ learning experiences. The focus of this review is on the practice of evaluating the performance of the provider, i.e. evaluating teaching quality and the various alternatives available.
A number of articles are assessed that deal with the evaluation of teaching and the need for such measures to address the students’ learning experiences. The paper also reviews experiences in validating the measures of teaching effectiveness, the sources of influence on the effectiveness of such measures, and what needs to be done to make the best use of such measures to improve student learning. Attempts are also made to relate the concepts discussed to ongoing educational quality improvement exercises in Ethiopian Higher Education Institutions.

One of the problems about evaluation of teaching is the lack of clear definition of what constitutes effective teaching. Even when attempts are made at defining the construct, there is the tendency to focus on what the teacher does in the classroom, implying that it is up to the teacher to prepare and provide knowledge to be grasped by the students. Teaching should not, however, be confined to what the teacher does in the classroom. Teaching effectiveness must also measure whether an appropriate environment is created where students are able to learn. Some authorities suggest, it is what the students can do, or what they believe they can do, that should be considered as indicators of the effectiveness of teaching (Ramsden, 2003). Another set of criteria is suggested by Chickering & Gamson (1987) who reviewed research works that described good teaching as one that encourages interaction between students and instructors, between students themselves, uses active learning, provides prompt feedback, emphasizes time on task, and respects diversity, talents and ways of learning. Despite lack of agreement on what constitutes effective teaching, it has long been suggested that teaching is effective when it results in student learning. Since the ultimate purpose of teaching is student learning, there can be no better criteria than those which assess changes in students as a result of the teaching.

It is amidst the debate about what constitutes effective teaching that this review on the evaluation of teaching is conducted. What alternatives are available to evaluate the effectiveness of teaching? How do we know the appropriateness of the measures employed? What other issues need to be considered in the evaluation of teaching quality? Below is a description of the alternatives employed.

**Student ratings of the effectiveness of teacher characteristics**

One of the approaches commonly employed to evaluate teaching is by considering students’ judgments of the effectiveness of certain teacher characteristics. This is because students have been directly involved in the instructional process and can provide appropriate comments for instructional improvement. The emphasis in most rating forms is, however, more on evaluating what the instructor does instead of what the students have learned as a result of the teaching. The effectiveness of the teacher characteristics is considered as indicator of the
quality of teaching, although several studies reveal that students anyway rate the educational experiences based on their perceptions of how much they learned in a course (Cohen, 1981; Marsh, 1987, 1991; Ryan & Harrison, 1995).

Students’ course grades and other performance measures
Learning gain, as indicated in the grades students score, is used as a measure of the effectiveness of teaching. Studies indicate that there is an association between grades and teacher ratings since student rating is higher in courses where student achievement is higher (Baird, 1987; Cohen, 1981). However, other extraneous factors such as student characteristics (intelligence, attitude, and socioeconomic status), the learning environment (class size, availability of learning resources), and the nature of the tests could influence the outcomes (Berk, 1988). What is more, in training programs where no standards are set, course grades are assigned based on relative performance and may not indicate what each student has actually achieved. In such instance, it is difficult to rely on course grades as indicators of student learning or the quality of teaching (Warren, 1992). The nature and types of assessment techniques used to determine student gains may also be varied and it is difficult to evaluate the effectiveness of the teaching amidst such differences (Ramsden, 1991). Moreover, the validity of the questions depends on the item writing skill of the teacher. It could also be the case that what is being tested may not be the taught curriculum and it is difficult to rely on the examination scores as authentic measures of student achievement (Biggs, 2003; Ramsden, 2003).

Other indicators of student outcomes include looking at student products such as written reports, presentations, models produced, or performance in simulated/laboratory environments. Although such performance outcomes are better indicators of learning, various teachers may use different criteria to evaluate such products. Differences in the skill of the observer, the method of evaluating the performance, and the criteria chosen to evaluate the performance may also influence the results. Hence, if student achievement is to be used as a direct measure of the effectiveness of teaching, then there should be a comprehensive statement of the objectives of the program, the range of performance expected to be accomplished, and an assessment of the pre-entry behavior of the students (Greenwood & Ramagli, 1980).
Students perceptions of their learning gain
Students make self-reflective judgments about teaching based on whether they have learned the contents of instruction. The measures of effective teaching should, therefore, be designed to assess the extent to which teaching and learning activities enhanced student learning (Nimmer & Stone, 1991; Ryan & Harrison, 1995; Ellet et al. 1997). Studies indicated that there is a better relationship between learning perception and rating than between course grades and rating (Arthur et al., 2003; Baird, 1987; van Os, 1999).

Considering students’ perceptions of their learning alone as an indicator of the effectiveness of instruction has certain limitations. It will be difficult to infer whether students’ perceived gain is the result of the effectiveness of the teaching or the outcome of their own efforts. Students may work harder to compensate for poor teaching and master the required knowledge or skill through own efforts (McKeachie, 1990). Moreover, perception of learning may not be the same as actual learning (Arthur et al., 2003). Hence, it is equally important to evaluate the quality of the instructional process to supplement the data on students’ perceptions of their learning.

From the descriptions above, it seems it is difficult to rely on a single approach to the evaluation of teaching quality. A number of factors are believed to influence student learning and only one source of information may not provide appropriate data on the effectiveness of teaching. The measure designed to evaluate teaching quality should therefore include achievement indicators expressed in the form of grades or other performance measures, students’ perceptions of their learning, as well as the effectiveness of certain teacher characteristics. Even so, it is also possible that the way the curriculum is organized can also impact on student learning. Hence, in addition to evaluating the teacher characteristics, students should also evaluation the appropriateness of the course for the intended purpose.

Ways of checking the appropriateness of the student ratings
It is one thing to have a system of evaluating the quality of teaching but another to prove that such a system is valid for the purpose. As part of the evaluation of teaching quality, it is mandatory for the higher learning institutions to present data to students, teachers and other stakeholder on the appropriateness of the measures employed. The following are the most commonly used approaches to check the validity of such measures.
Teaching improvement following the feedback

One of the purposes of conducting the evaluation of teaching is to provide teachers with feedback to improve their teaching. The changes observed among the instructors as a result of the feedback are seen as indicators of the appropriateness of the measures to assess teaching effectiveness (Tiberius et al., 1999; Lang & Kersting, 2006; van Os, 1999). Some research in this area shows that feedback has a profound impact on improving instructional practices, both during the ongoing semester (Hom et al., 1982) and even afterwards (Cohen, 1980; Wilson, 1986; Tiberius et al., 1989), which implies that the student ratings are valid. However, it is not always the case that comments from students help instructors improve their teaching, as the quality of the feedback depends on the appropriateness and design of the questionnaires. The feedback also has better impact if it is supplemented by discussion afterwards (Tiberius et al., 1989; Lang & Kersting, 2006) or when expert support is provided to instructors with lower ratings (Overall & Marsh, 1979; Wilson, 1986; Arubayi, 1987; Penny, 2004; van Os, 1999).

Although providing formative feedback from student ratings is considered as the most effective, there are problems due to lack of clear purpose for conducting the evaluation. Teaching improvement occurs when teachers are provided with detailed information on how to make these changes (Jacobs, 1987). Moreover, the way the instructors perceive the nature of the feedback influences its effectiveness. For instance, when there is no clear indication of why the evaluation is conducted, or if the instructors think that they have been criticized, they may not ask for any professional support (Moore and Koul, 2005). The evaluation of teaching is considered valid when the data suggests the required action for improvement and continuous support is provided following the feedback (van Os, 1999).

Relationship with other indicators of effectiveness

A number of studies have been conducted to assess the appropriateness of the student ratings as measures of effective teaching. A strong relationship with other indicators of effective teaching such as instructors’ evaluation of their own teaching and the similarity in perception of effective teaching held by students and instructors are some of the strategies used to check the validity of student ratings (Marsh & Roche, 1997). Although there are some concerns about the validity of student ratings, many studies have indicated that the student evaluation of teaching is a valid indicator of instructional quality since the ratings correlated positively with objective measures of student learning (Cohen, 1981; Marsh, 1984). More support for the validity of student evaluation of teaching comes from Cohen’s (1981) meta-analysis of multi-section studies that revealed
consistent relationship between achievement and the measures of teaching effectiveness employed. Moreover, there is evidence that over 90 percent of the explained variation in course ratings was associated with attributes that measure how much students learned in a course (Broder & Dorfman, 1994).

### The impact of conceptions of teaching and learning on the evaluation of teaching

The meaning teachers and students attach to teaching and learning has an impact on their judgments about quality. Teachers who conceive of learning as information accumulation to meet external demands, as transmitting information to students approach their teaching with teacher-focused strategies. On the other hand, teachers who conceive of learning as developing and changing students view teaching in terms of helping student learn. This raises the question of how these different conceptions affect the evaluation of teaching.

Research by Trigwell, Prosser, & Waterhouse (1999) indicated that there is a relationship between the teacher’s approach to teaching and the student’s approach to learning. Teachers describing their teaching as teacher-focused are more likely to have students reporting a surface approach to learning, while those teachers who focus more on the students and their learning are less likely to have students adopting a surface approach. What is interesting to note is that deeper approaches to learning are related to higher quality learning outcomes and to students’ perceptions of high quality teaching, independence, and awareness of goals and standards (Trigwell & Prosser, 1991). When teachers encourage the facilitation of learning, it is observed that their students apply a deep approach to learning (Kember, 1997).

Students who are actively engaged in learning for deeper understanding are likely to learn more than students not so engaged. When students apply a deep approach, they look for meaning in the material being learned and relate it with their previous knowledge. On the contrary, when students apply a surface approach to learning, they work harder to remember the material to reproduce it during examinations. Students, who conceived learning either as increase in knowledge or memorization employ a surface approach while learning a text. On the contrary, those who conceive learning as abstraction of meaning or understanding concepts apply a deep approach while learning the given text (Van Russom & Shenk, 1984).

There is also a link between conceptions and academic ability; i.e. students with transformative conceptions of learning, as opposed to reproducing, are observed to have higher academic standing (McLean, 2001). Moreover, Akerlind’s (2004)
review indicates that teacher-centered orientation is less likely to produce high quality learning outcomes among students. In contrast, when teachers adopt the facilitative or the conceptual change approach to teaching, there is a better possibility for high quality student learning (Trigwell & Posser, 2004).

Implications for the evaluation of teaching

An assessment of teachers’ and students’ conceptions of teaching and learning is important when designing measures of teaching quality. There are various reasons for this. Firstly, as indicated already, there is a link between conceptions of learning and student learning (e.g. Trigwell & Prosser, 1996.). Secondly, if students and teachers have differing conceptions of the meaning of effective teaching, it will be difficult to make use of the data from student evaluation for teaching improvement. For instance, if questionnaire items are written in terms of what the teacher should do, students might assume that it is solely up to the teacher to provide input. The implication is that the students may develop a wrong conception of effective teaching and student learning. Although good teaching is one that promotes understanding, students who have been using such ratings may perceive good teaching as reproducing information and may not give value to teaching that emphasized understanding (Entwistle & Tait, 1990).

Kember & Wong (2000) revealed that students with passive beliefs about teaching and learning state qualities expected of teachers that most commonly appear in teaching evaluation questionnaires. Examples include the need for the teacher to be organized, communicate clearly, and state clear objectives. On the other hand, students with active belief about learning expected self-directed learning, role play, collaborative learning and project-based learning, qualities that are not commonly indicated in standard questionnaires. The task of the lecturer is understood implicitly as providing ready made material to be absorbed by passive students. The paper concluded that the standard questionnaires commonly employed in most universities cannot be used to evaluate innovative or student-centred teaching since they are designed based on teacher centred approach. There are, of course, attempts made to include certain dimensions that refer to learner-centred instruction in questionnaires designed based on the teacher-centred orientation. An example is Marsh’s (1987) Student Evaluation of Educational Quality (SEEQ). The problem, however, is since students’ conceptions of learning were observed to be related to their descriptions of teaching (Kember & Gow, 1994) those who conceive learning as an active process may not give high ratings to teaching which is teacher centred. On the other hand, when the students have a passive conception of learning, they are expected to be biased against teaching which requires active engagement.
Changing conceptions of teaching and learning

From the research results stated above it is possible to infer that students learn better when teachers employ the learning facilitation orientations to teaching and students apply the deep approach to learning. Hence, it is mandatory to make a shift in conceptions of teaching, i.e. from the transmitting to the facilitative approach (Kember & Kwan, 2000). Moreover, exemplary teachers are observed to be applying a deep approach while teaching (Andrews et al., 1996) and using various approaches to assist student learning (Dunkin & Precians, 1992).

A study by Ho et al. (2001) reveals that it is possible for teachers to change their conceptions through in-service training. After a critical examination of their existing beliefs and practices, the teachers were provided with training on better ways of approaching their teaching. Teachers received higher ratings after participating in the training while students taught by these teachers also showed improvements in their approaches to study. The study stressed the need for staff development programs to make a shift from focusing on providing teaching skills training to changing teachers’ conceptions of teaching to improve teaching quality.

Other issues to be considered

Purpose of conducting the evaluation

It is common for university documents to promote academic program evaluation and to advocate remedial action on evaluation outcomes. Quality is assured when there are clear linkages between institutional goals, learning and teaching activities, and appropriate systems for evaluating and rewarding teaching. Clarity is first necessary about the purpose of the evaluation, i.e. whether the results will be used for personnel decisions (summative evaluation) or as feedback for teaching improvement (formative evaluation).

Formative evaluation is conducted so that instructors can use the outcomes for teaching improvement. On the other hand, summative evaluation is intended to make administrative decisions on promotion and pay rises. Although some researchers (e.g. Marsh, 1984) argue that data collected using a student rating forms can be used both for promotion purposes as well as for providing feedback, some feedback for the instructor may not be appropriate for administrative decisions (Miller, 1988). What is more, teachers give more credibility to student evaluation of teaching if the data is to be used for professional and instructional improvement than for administrative decisions (Joshua & Bassey, 2004).
Using the outcomes of the evaluation
It is not conducting the evaluation per se that matters. Equally, if not more important, is what the teacher does with the outcomes of the evaluation. Harvey (2001) criticized higher learning institutions for collecting information from students, but failing to state clearly how such information is to be used for improvement purposes. Hence, it is essential to put in place a clear structure of responsibilities for conducting the evaluation, reporting the results, taking remedial action, and communicating the outcomes to those who provided the data.

Quality improvement efforts in Ethiopian higher learning institutions
Until recently, there were no clearly stated standards of quality in the higher learning institutions in Ethiopia. However, when the Higher Education Proclamation was issued in 2003 (Proclamation No. 351/2003), relevant centers have been constituted with clearly stated mandates of introducing quality culture and providing required guidelines in these institutions. The Higher Education Relevance and Quality Agency (HERQA) is one of these centers that has been designing draft guidelines and conducting the first ever external evaluation of quality in the higher learning institutions in the country. There was also a lot of support provided by the NUFFIC sponsored EQUIP project that has instituted the Academic Development and Resource Centers (ADRCs) in the nine older universities.

Before the establishment of the ADRCs there were limited quality improvement exercises in the various higher learning institutions. Staff development activities were not properly conducted or were provided as a one-time experience only when there was a perceived need. The nine old universities have now training manuals on instructional skills, course design and review, and ICT in education. Moreover, as one of the objectives of the ADRC’s was to develop quality assurance instruments and guidelines for the evaluation of individual courses and programs, quality care policy is drafted and required instruments such as the teaching evaluation form designed with the intention to link the assessment of teaching with the support to be provided under the staff development programs.

In relation to the evaluation of teaching quality the ADRCs have been instrumental in designing improved measure of teaching effectiveness. In the previous teaching evaluation form, more emphasis was given to evaluating the effectiveness of teacher characteristics and there were no clearly stated dimensions indicated. Whereas in the improved evaluation form the items are
put into certain categories of teaching effectiveness, although most of them still reflect the effectiveness of teacher characteristics. Moreover, an additional dimension is considered specifying instructional contexts such as laboratory, workshop and clinical settings.

There are, however, certain issues that need to be considered to refine the newly designed measure of teaching effectiveness. Firstly, at the level of the institutions the purpose of conducting the evaluation is not clearly stated, although implicitly the outcomes are used for summative purposes when promoting instructors. In such instances, it is difficult to relate the data from student evaluation to the staff development programs provided at the ADRCs. The higher learning institutions should, therefore, state the need for the measures of teaching quality to focus on rewarding teaching that facilitate student learning and integrating data from student evaluation of teaching with staff development programs.

The most commonly used measures of teaching effectiveness, such as the Student Evaluation of Educational Quality (SEEQ) (Marsh, 1982) are designed based on one or a combination of the following approaches: asking students to describe indicators of effective teachers; providing a list of characteristics and requesting students to select appropriate measures of teaching effectiveness; and looking at the characteristics of outstanding teachers. Once the items are developed, they are put into possible dimensions of teaching effectiveness using appropriate statistical procedures, such as principal components analysis. The items in the improved measure of teaching quality are, however, designed by teachers, although students in some institutions were requested to give their comments during the tryout. Hence, there is still room for improvement in terms of considering students’ experiences and designing the items using appropriate procedures.

Universities these days are urged to be accountable to student learning. As a result, there is a shift of emphasis from identifying teacher characteristics as indicators of effectiveness to looking for the components of effective learning. Unfortunately, the items in the improved measure of teaching quality are still designed to reflect to the characteristics of effective teachers. Students’ evaluation of their learning can be considered as an additional source of information on the effectiveness of teaching. What is more, there is a lot more to do in terms of assessing teaching and learning conceptions of teachers and students in such a way that teacher focused orientations can be changed and the items in student evaluation forms can be redesigned to reflect student learning.
In conclusion, it is essential for higher learning institutions to put in place quality enhancement mechanisms as a way of improving the quality of their programs, one of which could be a system of evaluating the quality of teaching. Existing measures of teaching quality give more emphasis to the evaluation of the effectiveness of teacher characteristics as proxy indicator of student learning. Moreover, although there are various sources believed to influence student learning, they are not considered in designing such measures. Hence, it is essential for the measures of teaching quality to include as many sources of data on teaching quality, such as indicators of student learning, teaching practice, as well as course related elements, since it will be possible to provide comprehensive information about the effectiveness of teaching. For instance, Marsh’s SEEQ has one dimension called ‘learning value’ that refers to students’ interpretations of what they have learned in the course in addition to other dimensions that refer to teacher and course characteristics. A related issue that needs to be addressed is that such measures should specifically state the purpose of conducting the evaluation and make an assessment of the existing conceptions of teaching and learning among teachers and students while designing such measures.

References


