

Ethicality in EFL Classroom Assessment: Bridging the Gap between Theory and Practice

Shaimaa A. Torky and Nawal Abdul Sayed Haider

Abstract In recent years, the renewed focus on classroom or assessment for learning and the role it plays in gauging and supporting the learning process has entailed adopting an alternative approach to fairness that takes into consideration the open ended, dynamic nature and informality of most of these practices. The current study tackles EFL university teachers' perception of the ethicality of various classroom assessment practices to uncover the hidden code of ethics they ideally refer to, and to determine its conformity to codes endorsed by previous research. A sample of 28 teachers from the English department at the Public Authority of Applied Education (PAAE) at Kuwait University, College of Science, was selected. A survey in the form of a fifty-item questionnaire comprising five dimensions was utilized to assess teachers' assessment practices and their perceptions of assessment ethicality. Results imply that many areas were considered controversial for most teachers. One of these areas is using multiple forms for assessing students. Another issue is consistency between the assessment methods used and the curriculum objective and classroom activities. Equity issues also seem to be blurred for most teachers. Results of the current study testify to the value of training that is particularly focused on fair assessment and ethicality dilemma.

Keywords Classroom assessment · Assessment ethicality · Assessment for learning · Ethical conflicts · Teachers' perceptions of assessment

S.A. Torky (✉)

The National Centre for Educational Research, Cairo, Egypt

e-mail: shaimaatorky@gmail.com

N.A. Sayed Haider

Public Authority for Applied Education and Training, Adailiyah, Kuwait

e-mail: nawala33@hotmail.com

© Springer International Publishing Switzerland 2017

S. Hidri and C. Coombe (eds.), *Evaluation in Foreign Language Education in the Middle East and North Africa*, Second Language Learning and Teaching, DOI 10.1007/978-3-319-43234-2_4

1 Introduction

The concept of assessment has undergone a recognizable shift throughout educational history. Originally, assessment was conceptualized as testing learners to provide evidence of accountability for stakeholders. In fact, this view is thought to ignore the fact that assessment should be geared first to assist students and provide them with the necessary feedback regarding their progress and efficiency of their adopted strategies (Buzzelli & Johnston, 2002). Most scholars consider this approach as “assessment of learning” (A \mathcal{O} L).

Yet, with the advent of concepts like lifelong learning, social construction of knowledge and learner’s autonomy, the term assessment has evolved to endorse an alternative conceptualization that views assessment as the most important means of providing help to learners and fostering their self-awareness. Akin to this paradigm shift, unconventional means of assessment that drastically change power relationships in the classroom have emerged; the call for using self-assessment, peer assessment and participatory assessment is a clear reflection of a more democratic or participatory trend (Tierney, 2010). This quite recent notion is termed formative assessment or according to Dann (2002) and Tierney (2010) assessment for learning (A \mathcal{F} L).

However, although teachers may spend as much as one third of their time in assessment-related activities (Stiggins, 2002), pre-service and in-service training in Kuwait do not require EFL teachers to take a course, or demonstrate competency in the area of assessment, suggesting that teachers often lack formal training in assessment. Even if some training is provided in this regard, it is almost tackling the procedures of designing and applying assessment tools. Hardly any guidance is provided that touches on ethical dilemmas EFL teachers might encounter as they attempt to strike a balance between the two competing demands of tuning into students’ needs on the one hand, and meeting the demands for accountability, on the other hand. This lack of preparation in assessment related ethical issues is problematic because ethical reasoning in assessment can barely develop by everyday experience (Green, Johnson, Kim, & Pope, 2007). Hence, left to their own devices, teachers’ decisions or classroom practices pertinent to preparing, designing tests and grading students’ on various assessment tools are mostly based on intuition.

2 Literature Review

Classroom assessment is small-scale assessment prepared and implemented by teachers in classrooms. Classroom assessment includes traditional assessment and alternative assessment. Traditional assessment refers to predetermined testing measures such as selected-response tests (e.g., multiple-choice questions, true/false questions, matching questions), brief constructed-response tests (e.g., short-answer questions), and essay questions. Alternative assessment refers to authentic

assessment tasks/forms such as oral questioning, teacher observation, performance tasks, and student self-assessment (Airasian, 2005; McMillan, 2007).

Traditionally, a fair test was considered one, which is free from bias, partiality, discrimination and favouritism (Tierney, 2010). Adopting this point of view, the Educational Testing Service (ETS) (2002) has added a new section addressing the issues of eliminating bias and ensuring equity in the testing processes. Notably, the fairness concept adopted by these standards is highly technical, and so it was thought that evidence of fairness could be obtained via statistical procedures such as validity and reliability (Camili, 2006; Volante, 2006). From another perspective, a broader procedural conceptualization for fairness that goes beyond statistics has started to emerge. This includes defining a clear purpose for the test, developing test specs, reviewing test content and, finally, conducting a field test of the examination (Plake & Jones, 2002). Noticeably, ethical issues are not directly tackled according to this approach.

Attempting to address the ethical issues of assessment more deeply, Messick (1995) stressed the interrelatedness of ethicality and validity. According to his viewpoint, test developers need to minimize construct-irrelevant test variance, which can be the result of the test response being based on factors irrelevant to the objectives being assessed and thus might distort results. Fair testing has also become a pursuit for standardized testing practices (JCSEE, 2003, p. 3). Fairness is, also, a critical consideration for good testing practice in ILTA's (International Language Testing Association) Draft Code of Practice (2000). Noteworthy, however, this discussion of fair testing is very limited and disregards ethical assumptions underlying day-to-day classroom assessment.

Thus, in recent years, the renewed focus on assessment for learning has entailed adopting an alternative approach to fairness that reemphasizes the value-laden aspects that have long been neglected in education (Blanchard, 2008). Providing an operational concept of fairness that takes into consideration the unique aspects characterizing assessment for learning, Airasain (2005), Camili (2006), McMillian (2007), Shepard (2005, 2007) and Zhang and Burry-Stock (2003) argued that fairness includes setting clear learning expectations, helping students learn how to do the assessment task, ensuring equity and avoiding bias, using varied approaches for eliciting learning, accommodating special needs and providing detailed and balanced feedback for learners.

This approach to fairness, as seen from the aforementioned criteria, places the learner and his interests at the heart of the assessment procedure. Therefore, great emphasis is placed on preparing the learner for the assessment. Noticeably the assessment is no longer viewed as an end in itself, rather it is seen as a tool to drive the learning process forward. Nonetheless, Tierney (2010, p. 63) views these standards as lacking sufficient description or empirical evidence that attest to their validity. Also missing, according to his viewpoint, is the emphasis on the protection of privacy, communication of results and the use of multiple evaluators.

From another perspective, the concept of fairness can be quite related to the ethical dilemma teachers face in their relationships with the individuals they interact

with in their professional life. Reviewing previous research, some codes could be gleaned, such as the no harm principle, avoiding score pollution, equity, transparency and consistency. In fact, all these codes are intertwined so that it renders it difficult to draw clear broader lines to separate them.

The concept of no harm, unlike the abstract concept of fairness which may fall short of revealing teachers' practices, stimulates their awareness of their mal-practices regarding classroom assessment. Examples of harm can be unexpected items on or offensive items in a language test. Avoiding score pollution is an application of the principle of "do no harm" to assessment. It is defined as any practice that improves test performance without concurrently increasing actual mastery of the content tested (Payne, 2003). When teachers take into consideration during grading students' work are factors, such as their effort, behaviours or punctuality, the scores students obtain may overstate or understate their actual skills. Inequity, as opposed to equity, occurs when the teacher offers different opportunities and makes different decisions in the same environment. For instance, teachers sometimes unjustifiably increase test time for some students or change students' answers (Gipps & Murphy, 1994; Hidri, 2015). Transparency and consistency can also affect test fairness. Transparency implies involving students in the process of determining the evaluation criteria and methods of assessment (Popham, 2000). Consistency basically necessitates compatibility between assessment tools and the purpose for which they were designed (Pope, Green, Johnson, & Mitchell, 2009).

2.1 Teachers' Perceptions of Assessment

As Chang (2006) argues, few studies attempted to probe into their underlying perspective and ethical convictions that inform their various decisions. Nevertheless, examining previous studies, two research trends could be discerned: The first aimed at investigating teachers' perspectives of their assessment practices in general; the second, however, focused on examining teachers' ethical convictions underlying their practices.

Addressing the first trend, Pelly and Allison (2000) explored primary school teachers' perspectives on the assessment of the use of the English language in Singapore. The findings revealed that teachers were markedly divided and uncertain in their views of the efficacy of current tests. Zhang and Burry-Stock (2003) investigated teachers' assessment practices across teaching levels and content areas. Results showed that, regardless of teaching experience, teachers with measurement training reported a higher level of self-perceived assessment skills than those without measurement training. In the same way, Chan (2008) investigated elementary teachers' beliefs and practices of multiple assessments. Results showed that most teachers considered using multiple assessments a positive experience. Yip and Cheung (2005) pinpointed that teachers expressed their concerns about the

consistency of assessment whether among different teachers or consistency in each individual teacher's assessment practices.

Two studies used the Teachers' Conceptions of Assessment (TCoA) inventory to investigate teachers' assessment conceptions. The first is Hidri's study (2015) and it revealed that teachers harbour wrong and conflicting assessment conceptions. The second is Gebril and Brown's study (2013), which suggested that greater changes to the examination system are warranted if teacher beliefs are expected to be more positive about the priority of formative, improvement-oriented uses of assessment.

To address the second trend, some researchers attempted to examine teachers' ethical beliefs and the ethical dilemmas they grapple with during everyday classroom practices. For instance, Szyrka (2001) explored the relationships between equitable assessment practices and actual classroom assessment practices. Results showed discrepancies between teachers; whereas some teachers believe that tasks should be modified to enable each learner to be successful, others think that all students should abide by the same standards. Lu (2003) investigated the beliefs and practices of assessment by two university English instructors. The results showed that there was a high consistency and a very slight inconsistency between the instructors' beliefs and their assessment practices.

In the same way, employing a web-based survey, Green et al. (2007) conducted a study that exposed teachers to a set of classroom scenarios to examine their implicit ethical perspectives. Findings suggest that assessment is currently an educational realm without professional consensus. Likewise, Tierney (2010) and Simon, Chitpin, and Yahya (2010), found, throughout studies that aim at reaching a better understanding of classroom assessment fairness, that teachers' fairness relies on their ability to understand students and to reflect on both the interaction and decisions made in the classroom. Moreover, the studies revealed that group work, test failure, fairness, multiple assessment opportunities, and academic enablers were key areas of concern.

Pope et al., (2009) conducted a study to document ethical conflicts faced by teachers regarding the assessment of students. Critical incidents generated by teachers revealed a majority of reported conflicts related to score pollution, and conflicts frequently arose between teachers' perceptions of institutional demands and the needs of students. Using an introspective critical approach, Simon, Chitpin and Yahya (2010), examined pre-service teachers' perceptions of classroom assessment. The researchers found that group work, test failure, accommodation, fairness, multiple assessment opportunities, and academic enablers were key areas of concern for most teachers.

The only study that approached classroom assessment from the students' perspective was the study of Bursuck, Munk, and Olson (1999) who attempted to determine the students' perceptions about fairness of grading their final report. Students indicated that they could accept differentiation in teacher's responses to an assessment task, yet they cannot accept adaptation in the assessment task to accommodate various students' needs.

2.2 *Writing Assessment*

Being highly prone to teachers' personal judgment or to students' self-judgment, the accuracy and reliability, and hence fairness, of writing assessments are looked at with a great deal of skepticism. Fairness in writing assessment was tackled from two perspectives. On the one hand, teachers' perspectives on the fairness of the criteria they adapt to grade students' performance are tackled. On the other hand, students' self-assessment as a means of realizing equality is also probed.

Addressing the first perspective, Zoeckler (2005) intended to understand the moral aspect of grading writing by examining English language teachers' assessment artifacts and by interviewing and taking field notes. Similarly, Graham (2005) conducted a two-year study on pre-service teachers to track the development in their assessment theories and practices. Teachers agreed that evaluating writing is a challenging process and their comments regarding fairness centered on providing constructive feedback for weak students to provide them with the best possible support.

Similarly, Dann (2002) conducted a case study that aimed at examining the fairness of grading students' writing projects. A participatory form of classroom assessment was adopted where self, peer and teachers' assessments were embraced. Though most students expressed their overall satisfaction with the scores they obtained, they expressed a sense of having the scores imposed on them that shows that participatory classroom assessment can be controversial.

Therefore, it can be concluded from the previous literature review that, except for the study of Green et al. (2007) and Pope et al. (2009), few studies have investigated EFL teachers' perceptions of assessment ethicality or the ethical dilemmas that can bear upon their classroom practices. Noteworthy, also, research examining teachers' perceptions of assessment in general and ethical considerations pertinent to assessment in particular has either resorted to using indirect methods, such as interviews, scenario techniques and incident techniques or direct methods, e.g., direct classroom observation, artefacts or examining students' perceptions. Most of these studies have widely reflected the paradoxical stance most teachers experience when they assess students. Issues such as equity, fairness or score pollution and distinction between what constitutes fair assessment versus unfair assessment are still blurred for most teachers. Furthermore, the concept of assessment for learning with its implications has not yet been well assimilated by teachers, and hence their notion of equity is rather confined to the traditional concept of assessment of learning. Thus, the purpose of the current study is to lay more emphasis on the concept of ethicality and how teachers perceive this concept as far as their assessment practices are concerned.

3 Purpose of the Study

The purpose of the current study is twofold. On the one hand, it is meant to examine EFL university teachers' perceptions of the ethicality of various classroom assessment practices to uncover the hidden code of ethics they ideally adhere to, and to determine its conformity to codes endorsed by previous research. On the other hand, the study aims at examining EFL teachers' current practices regarding various ethical issues in the realm of classroom assessment to identify the discrepancy between those practices and teachers' ethical beliefs. The study aimed at answering the following questions:

- (a) What are EFL teachers' perceptions regarding the ethicality of the identified classroom assessment practices?
- (b) What are the actual assessment practices carried out by those teachers in light of ethicality considerations?
- (c) To what extent are teachers' perceptions and classroom assessment practices consistent with ethicality norms?

4 Method

The current study used both qualitative and quantitative methods to investigate teachers' perceptions of ethicality of their classroom assessment practices. The perspectives of the participants are of the utmost importance as the researchers sought to understand and describe the participants' experiences as seen by the participants themselves. For these reasons, therefore, the application of the qualitative paradigm was considered critical to the study. This took the form of analysing teachers' answers on each questionnaire item by item to discern their way of thinking and locate compatibility, or lack of it, between their endorsed ethical codes and classroom practices. Quantitative methods were utilized, however, to analyse obtained data and get generalizations about teachers' perspectives.

4.1 Participants

Purposive sampling was utilized to locate teachers who were willing to converse about their experiences with classroom assessment practices. A sample of 28 teachers, 16 females and 12 males, from the English department at the Public Authority of Applied Education (PAAE) at Kuwait University, College of Science, was selected. Respondents had taught for an average of fifteen years. Current grade level taught was university levels. Some respondents (30 %) had a bachelor's degree; (25 %) held a Master's degree and (45 %) held a Ph.D. About 82 % of the

teachers had had at least one measurement course. Teachers were involved in a multitude of assessment activities including administering regular exams, grading writing, evaluating students' oral performance and applying final university mandated exams. Thus, it was thought that the sample of this study would have adequate experience or background knowledge of classroom assessment, and so they would be able to judge the ethicality or otherwise of various practices. First, Tables 1, 2 and 3 present summary information on respondents by teaching experience, assessment experience and obtained training.

It is clear from Table 1 that the teachers' teaching experience ranged from less than 7 to more than 22 years, with most teachers having less than 7 years of experience. As Table 2 shows, teachers' experience in assessment ranged from 1 to 20 years, with most teachers having from 16–20 years of experience. Table 3 indicates that about half of the teachers (46.7 %) had not received any training courses in assessment, (21.7 %) received pre-service training and (10.7 %) received in service training. Yet, (21.4 %) reported that they received training throughout other means such as conferences, individual reading and workshops.

Table 1 Teaching experience

EFL experience (in years)	Number	Percent	Cumulative percent
<7	13	46.7	46.7
15–21	7	25	68.4
22–	8	28.6	100
Total	27	100	

Table 2 Assessment experience

Experience (in years)	Number	Percent	Cumulative percent
1–10	10	35.7	35.7
11–15	7	25	57.1
16–20	11	39.2	100
Total	28	100	

Table 3 Professional assessment related training

Assessment training	Number	Percent	Cumulative percent
No training courses	13	46.7	46.7
Pre-service training	6	21.4	67.9
In service courses	3	10.7	78.6
Other means of training	6	21.4	100
Total	28	100	

4.2 Instruments and Procedure

A survey in the form of a questionnaire, consisting of 50 items, comprising five dimensions, was utilized to assess teachers' assessment practices and their perceptions regarding assessment ethicality. The instrument was developed within the theoretical framework delineated by the literature on classroom assessment and fair testing.

Teachers were asked to mark their responses to the same 50 items on two different rating scales: The practice scale and the ethicality scale. The practice scale was designed to measure teachers' assessment practices on a 3-point scale (*1 = never practiced, 2 = sometimes practiced and 3 = usually practiced*). The ethicality scale was designed to measure teachers' ethicality perceptions on a three-point scale (*1 = unethical, 2 = somewhat ethical and 3 = ethical*). Negatively-keyed items were "reverse-scored" before computing students' total scores. These included all items subsumed under the dimension of score pollution as well other items with the sign (-) as shown in the Appendix. Two data sets were produced, one on assessment practices and the other on perceived ethicality assessment skills. The items of the questionnaire are presented in the Appendix.

The questionnaire comprised five dimensions reflecting ethicality in assessment. Though an overlap in the underlying dimensions may exist, each dimension contains a certain degree of uniqueness. The first dimension is transparency and confidentiality; it subsumed 5 items addressing the teachers' openness regarding assessment objectives, techniques and correction methods. It also subsumed two items tackling the examinees' right for privacy. The second dimension comprised 8 items- and it implied making sure that the assessment adheres to its purpose and to what was taught in the classroom. The third dimension is avoiding score pollution, and it comprised 12 items measuring teachers' awareness of the fact that the score a student receives should tightly reflect what he/she mastered. The fourth dimension is A/L-comprising 8 items. This dimension covered aspects, such as using peer evaluation, using multiple assessment methods and avoiding looking at testing as the sole high-stakes assessment device. The last dimension addressed in this questionnaire is equity-including 17 items-, which addressed bias avoidance, providing equal chances to students and catering for various students' needs.

The scenario technique was adapted to present teachers with a set of classroom assessment situations reflecting various stances of classroom assessment, including various sorts of formal and alternative classroom assessment. The assessment situations tackled were derived from everyday classroom experience and were categorized under a set of main areas: Preparation for assessment, developing assessment tools, administering assessment, grading and feedback and reporting or communicating grades.

The first draft of the survey consisted of 62 scenarios and 6 questions about demographic information. To establish validity of the questionnaire, selected professors from the field of assessment and experienced EFL teachers in Kuwait University were asked to review the survey questions. Some items were deleted and

others were modified according to the jury viewpoint. Subsequently, a pilot survey was conducted with 14 participants. Participants were given oral instructions on how to answer questions. The results of the pilot survey were reviewed and six items that appeared confusing were modified or replaced. Reliability analysis yielded a Cronbach's of 0.75, for the survey items. Reliability of all dimensions ranged from 0.55 to 0.72. This proves the consistency of the survey and of its various dimensions. The final 50-item survey was administered in the summer of 2013. The instrument along with a cover letter was distributed to the teachers by both researchers, while some were sent via email.

5 Results

5.1 Teachers' Perceptions and Practices

Determining teachers' practices of ethical assessment was based on the respondents' scores on the practice scale of a questionnaire containing 50 items. Similarly, teachers' perceptions of ethicality were based on their scores on the ethicality scale of the same questionnaire. The scores were a sum of these 50 items.

Frequencies and percentages were used to summarize teachers' rating of each situation in terms of the frequency of practice as well as in terms of ethicality. From these percentages, implications could be drawn about ethical issues that were controversial. Moreover, teachers' malpractices or misconceptions regarding these issues were also pinpointed to identify areas warranting more focus. First, to analyse teachers' scores on the survey, descriptive statistics were obtained. Furthermore, to test whether EFL teachers' perceptions of multiple assessments were related to their practices, Pearson product-moment correlation coefficient was computed as shown in Table 4.

Table 4 shows that on the practice scale, teachers' scores ranged from a low of 81 to a high of 112. On the perception of ethicality dimension, teachers' scores ranged from a low of 100 to a high of 122. Teachers' overall mean score on the practice scale was (96.5) and the SD was (5.3). On the other hand, the overall mean score on the ethicality scale was (110) and the SD was (4.7). Given that the total score was 150, the mean score shows that teachers' practices can hardly be considered fair or ethical, even if their perception of ethicality shows that they were

Table 4 Teachers' practices and perception of assessment ethicality

	N	Total	Minimum	Maximum	Range	Mean	SD	Pearson correlation	Sig. (2-tailed)
Total ethicality	28	150	100	122	22	110	4.787	0.219	0.433
Total practice	28	150	81	112	31	96.5	5.364		

relatively aware of what constitutes ethical versus unethical assessment practices. The Pearson correlation was computed between EFL teachers' practices and perception practices, yielding a value of 0.219. The result showed that the relationship between beliefs and practices was not significant, $p = 0.433$.

Considering the scale sub-dimensions, further insights could be drawn. First, a Pearson product-moment correlation coefficient was computed to assess the relationship between teachers' practices and perspectives regarding transparency and confidentiality.

Teachers' mean scores on the practice aspect of the transparency and confidentiality dimension was (12.7) and the SD was (1.2). Similarly, teachers mean score on the ethicality aspect of the same dimension was (10.8) and the SD was (2.1). Pearson product-moment correlation coefficient between the teachers' practices and perceptions of ethicality was not significant at 0.05, $r = 0.008$, $p = 0.996$ (Table 5).

Regarding teachers' responses to the sub-items subsumed under the first dimension, descriptive statistics were obtained. In particular, in terms of preparation for assessment, most of the teachers (68 %) assigned high ethical value to unveiling their grading schemes to the students; this was also reflected in the practice of (68 %) of the study sample. As for assessment development, (28.6 %) of the teachers seemed unconvinced with the importance of sharing with students the rubric according to which a written task will be corrected, whereas (46.4 %) were unsure of whether to consider this practice as ethical or unethical. This uncertainty was also reflected in teachers' practice, i.e., only (39.3 %) indicated that they would entirely avoid hiding information about the writing rubric.

As far as the communication of results is concerned, only (32.2 %) of the teachers recognized the unethicality of limiting feedback to students' strengths; yet the majority (46.7 %) were unsure how to categorize such a practice. When it comes to practice, most of the teachers (53.7 %) reported that they would limit their feedback to students' strengths, whereas (39.3 %) would totally avoid that. The percentages of teachers who admitted to the unethicality of disclosing students' scores to their partners or to other parties were (35.7 %) to (42.9 %) for both cases respectively; however, in everyday practice, the majority of teachers agreed that they would never announce students' scores in front of their partners (67.9 %) or disclose a student's academic information to their peers (89.3 %).

As far as consistency is considered, descriptive statistics were obtained and a Pearson product-moment correlation coefficient was computed to assess the relationship between teacher's practices and perspectives.

Table 5 Teachers' practices and perceptions of transparency and confidentiality

	N	Minimum	Maximum	Range	Mean	SD	Pearson correlation	Sig. (2-tailed)
Practice	28	11	15	4	12.7	1.2	0.008	0.996
Ethicality	28	7	15	8	10.8	2.1		

Note Correlation is significant at the 0.01 levels

Table 6 Teachers' practices and perceptions of assessment consistency

	N	Minimum	Maximum	Mean	SD	Pearson correlation	Sig. (2-tailed)
Practice	28	15.00	22.00	19.11	1.99927	0.527	0.018
Ethicality	28	15.00	22.00	18.066	2.12020		

Table 6 shows that teachers' mean score on the practice aspect of the dimension of consistency was (19.1) and the SD was (1.99). Similarly, teachers' mean score on the ethicality aspect of the same dimension was (18.06) and the SD was (2.12). Since total score on this dimension is 24, it can be concluded that teachers adopted quite fair practices in terms of the conformity of the assessment utilized to both the purpose of assessment and the teaching methods adopted. Pearson product-moment correlation coefficient between teachers' practices and perceptions of ethicality was significant at 0.05, ($r = 0.53$), ($p = 0.018$). Since the coefficient of determination (r^2) = 0.28, the correlation between both constructs is considered low to moderate.

Analysing sub-items subsumed under this dimension, it appeared that, in terms of preparation for assessment, although most of the teachers (68 %) assigned high ethical value to practices pertinent to consistency, such as training students on test taking skills and administering a parallel form of the test, only (26.7 %) reported that they would usually administer a parallel form of the test.

With regard to developing assessment tools, (57 %) believed that any test has to be designed with reference to the curriculum objectives; this conviction was also reflected in the practice of (68 %) of the respondents. Similarly, (53.3 %) avoided incorporating methods that students have not encountered before, and (68 %) avoided using surprise items in their assessment. However, it seemed that teachers were quite unsure of the ethicality or otherwise of these practices; basically, only (46.7 %) thought that both practices are unethical. Similarly, the majority of teachers (80 %) reported that they usually try to incorporate assessment activities similar to those practiced in the classroom, which conformed to the beliefs of (80 %) of the study sample. When assessing oral proficiency, only (47.7 %) of the teachers would refrain from solely relying on classroom observation. These practices seem to conform to the teachers' perception of ethicality, i.e., only (26.7 %) perceived this practice as unethical; the rest were either unsure (46.7 %), or certain that classroom observation was sufficient to judge students' oral competence (13.3 %).

As far as grading is concerned, surprisingly, a high percentage of teachers (73.3 %) reported that they would sometimes grade reading comprehension based only on two multiple-choice tests. This practice is compatible with the ethical perspectives of all respondents since no teacher could perceive the unethicity of using a method that underrepresents students' competence.

Similarly, descriptive statistics were obtained and a Pearson product-moment correlation coefficient was computed to assess the relationship between teacher's practices and perspectives regarding the issue of avoiding score pollution.

Table 7 Teachers' practices and perceptions of score pollution

	Minimum	Maximum	Mean	SD	Pearson correlation	Sig. (2-tailed)
Practice	20.00	28.46	24.08	2.366		
Ethicality	17.00	25.00	22.38	2.471	0.172	0.054

As Table 7 shows, the mean score the teachers obtained on the practice aspect of "avoiding score pollution" dimension was (24.08) and the SD was (2.36). This mean is considered low relative to the total score which is 36. In the same vein, teachers' mean score on the ethicality aspect regarding score pollution was (25), which is low as well. It also shows that the teachers lack awareness of what constitutes ethical versus unethical assessment practices regarding score pollution. Pearson product-moment correlation coefficients between teachers' practices and perceptions of ethicality was not significant at 0.05, $r = 0.172$, $p = 0.541$.

Regarding teachers' responses to the sub-items subsumed under this dimension, frequencies and percentages of teachers' responses showed that, in terms of preparation for assessment, most of the teachers seemed uncertain as to the ethicality of various practices relevant to avoiding score pollution. In particular, some teachers (33.3 %) could not perceive the unethicality of practices aiming at pre-exposing students to parts of an upcoming test, while (47.6 %) were unsure or felt divided. Yet, in practice more teachers (46.7 %) reported that they would totally avoid training students on specific activities included in the actual assessment. Moreover, (46.7 %) of the teachers did not find it ethically problematic to draw students' attention to certain materials to prepare for an exam, while (33.3 %) were unsure about such a practice. Similarly, practice-wise, most teachers decided that they would draw students' attention to important material either on a regular basis (60 %) or sporadically (33.3 %). Notably, most teachers (93 %) agreed that in actual situations, they would not deduct more points for a wrong answer than for leaving the answer blank, though they were somewhat sceptical as to the ethicality of such practice.

As regards assessment development, providing clues to help students figure out the answer was considered unethical by (46.7 %) of the teachers which was also reflected in the practice of (40 %) of the respondents. Most teachers (66.7 %) found it totally unethical to pinpoint the correct answers through using a higher voice pitch, yet in practice only (40 %) reported that they would entirely refrain from alluding to the correct answer, whereas (53.3 %) decided that they would sometimes allude to correct answers. Notably, although (53.3 %) reported that they would refrain from vocally placing more emphasis on certain parts of test instruction to allude to the right answer, teachers seemed to have a blurred vision of what constitutes ethicality in that regard; only (33.3 %) referred to that practice as unethical, the rest were either unsure (33.3 %) or judged the practice as ethical (33.3 %).

As for grading and providing feedback, teachers seemed to hold a great deal of ethical misconception regarding the fairness of the grade students are awarded. All

teachers (100 %) considered it ethical to deduct scores for late work and to count students' effort or participation in the final grade. Teachers also seemed to have a blurred vision regarding whether to fail students for missing an exam (80 %). Others (80 %) believed it is ethical to grade students for exhibiting mastery even if class work was not completed. These beliefs were somehow reflected in the practices of (73 %) of the teachers who decided to count how late the homework was handed in, how much effort the student exerted (100 %), and in the case of group work, they would count other students' effort (86.6 %). In the same way (66.7 %) of the teachers reported that they would assign students a good mark for showing content mastery regardless of course assignment completion. Likewise, (46.7 %) of the teachers reported that ethicality-wise, a student who missed an exam should deserve the same score as that of a failing student. This was reflected in the practice of (60 %) of the respondents.

5.2 Assessment of Learning

To analyse students' scores, descriptive statistics were obtained and a Pearson product-moment correlation coefficient was computed to assess the relationship between teachers' practices and perspective regarding the issue of consistency. As Table 8 shows, the mean score the teachers obtained regarding their practice on the dimension of AfL was (15.9) and the SD was (2.1); this mean is considered low relative to the total score which is 24. The scores ranged from a low of 13 to a high of 19. In the same way, teachers' mean score on the ethicality aspect of AfL was (16.8), which is also quite low. The SD was (2.09). The scores ranged from a low of 14 to a high of 20. This gives indication to the fact that teachers normally do not adopt practices that reflect the use of assessment to enhance the learning process, which might allude to a lack of understanding of what, constitutes ethical versus unethical assessment practices. Pearson product-moment correlation coefficients between teachers' practices and perceptions of ethicality was not significant at 0.05, $r = 0.506$, $p = 0.054$.

Regarding teachers' responses to the sub-items subsumed under the fourth dimension (AfL), in terms of preparation, most of the teachers (86.7 %) seemed certain of the necessity of comprehensively covering the content before considering assessing students' mastery. This was also clearly reflected in the practice of the majority of the respondents (86.7 %) who agreed that they would not assess students until they had made sure they have covered the intended material. In the same way,

Table 8 Teachers' practice and perception of assessment for learning (AFL)

	N	Range	Minimum	Maximum	Mean	SD	Pearson correlation	Sig. (2-tailed)
Practice	28	6.00	13.00	19.00	15.9	2.144	0.506	0.054
Ethicality	28	6.00	14.00	20.00	16.8	2.09		

(73.3 %) believed in the ethicality of using multiple means for assessment; this was also reflected in the practice of (60 %) of the teachers.

When it comes to grading and providing feedback, some uncertainty as to whether to consider peer evaluation of oral performance an ethical practice could be observed, only (6.7 %) thought it is fair or ethical to include peer assessment in the students' final grade, whereas (73.3 %) were unsure or quite divided in their opinions; correspondingly only (13.3 %) agreed to regularly include that kind of rating in assessment. However, when it comes to writing assessment, (46.7 %) teachers seemed to be more tolerant to accept peer rating as an ethical practice. Nevertheless, a discrepancy can be discerned when examining teachers' practice. Only (13.3 %) of the teachers reported that they usually resort to peer rating to correct either oral reports or writing performance; others employed peer assessment sporadically—(23 %) and (33 %) for both oral and written performance respectively. This shows that teachers are still reluctant about involving students in the assessment process; that also explains why many teachers reported that they usually (46.7 %) or occasionally (40 %) weigh tests heavily compared to other means of assessment. This was backed by a strong conviction that testing should be given precedence; (86.6 %) of the teachers thought it is totally or somewhat ethical to weight tests heavily.

As regards to communicating students' results, considerably although student-teacher conferencing was considered ethical by (73.3 %), only (40 %) confirmed that they would regularly perform conferencing sessions, while (33.3 %) reported that they would use conferencing occasionally. Similarly, slowing down the teaching pace according to students' results was considered ethical by (66.7 %) of the teachers. This was reflected in the practices of (33.3 %) of the respondents, who pinpointed that they would act responsively to students' results, and also in the practice of (53.3 %) who indicated that they would occasionally adopt that responsive action. Categorizing students and labelling them as high, low, at risk was thought to be unethical by only (33.3 %) of the study sample, and was as well reflected in the responses of (40 %) of the teachers. The rest decided to label students according to their level either occasionally (46.7 %) or regularly (13.3 %).

To analyse students' scores on the dimension of equity, descriptive statistics were obtained and a Pearson product-moment correlation coefficient was computed to assess the relationship between teacher's practices and perspectives regarding equity. As indicated in Table 9 shows, teachers' mean score on the ethicality aspect of the "equity" dimension was (39.3) and the SD was (3.5); this mean is considered moderate relative to the total score which is 51. In the same way, teachers' mean score on the practice aspect of the "equity" dimension was (37.1), and the SD was (3.9). The scores ranged from a low of (27) to a high of (44). This gives indication to the fact that teachers somehow adopted assessment practices to ensure fairness and equity among students. Pearson product-moment correlation coefficients between teachers' practices and perceptions of ethicality was not significant at 0.05, $r = 0.262$, $p = 0.346$.

As for teachers' response to the sub-items subsumed under fifth dimension-equity, percentages and frequencies show that, in terms of assessment

Table 9 Teachers' practice and perception of equity

	N	Range	Minimum	Maximum	Mean	SD	Pearson correlation	Sig. (2-tailed)
Ethicality	28	10.00	37.00	47.00	39.3	3.514	-0.262	0.346
Practice	28	17.00	27.00	44.00	37.1	3.99		

development, most of the teachers (60 %) seemed certain that any test should cater for students' interests, yet the majority (60 %) occasionally reported that they would follow that practice. Providing help to weak students was considered unethical by (40 %) of the teachers, yet (46.7 %) were sceptical as to the ethicality of such practices. Nonetheless, as far as practice is concerned, (60 %) of the teachers reported that they would avoid giving extra clues to weak students. Taking into account the needs of students with special needs was considered ethical by (60.7 %), yet only (40 %) of the teachers indicated that they would attune, via assessment methods, to students' special needs.

As for administering assessment tools, (46.7 %) considered it ethically unproblematic to draw a student's attention to an item he has missed by mistake. Correspondingly, only (32.3 %) would regularly resort to that practice or would do so occasionally (32.3 %). This viewpoint regarding missed items was not endorsed when it comes to items students answered wrongly; most of the respondents (71.4 %) agreed that it was unethical to draw students' attention to incorrectly answered questions, yet in actual classroom situations only (32.3 %) would totally avoid correcting students' incorrect answers. In the same way, reminding students of what was learned during a test was considered unethical by (73.3 %). The practices of both translating difficult words during a test and giving slow students extra time were considered unfair by only (20 %) of the respondents. With regard to practice, (40 %) of the teachers reported that they would entirely avoid translating words, while only (32.3 %) would avoid giving slow students extra time.

In terms of providing feedback, teachers seemed to view bias towards weak students as an ethical practice, i.e., only few teachers admitted to the unethicality of giving extra marks to a weak class (33.3 %), however, a higher number of teachers (53.3 %) agreed on the unethicality of bias against an advanced class. As far as practice is concerned, only (40 %) would entirely refrain from being less strict with a weak class, whereas (53.3 %) of the teachers would not deprive high-level classes from getting chances for getting extra marks. Unexpectedly, teachers' stance towards the ethicality of relying on teachers' discretion in assigning grades was not quite definite; only (33.3 %) agreed that relying on the teacher's own impression is unethical. On the contrary, when queried about their practice, most teachers (53.3 %) decided that they would avoid unsubstantiated conclusions even if they sounded self-evident.

Addressing students' individual circumstances formulated a part of teachers' perceived ethical dilemma. For instance, although only (40 %) of the respondents thought it was unethical to be biased toward a student due to his unprivileged economic condition, about (60 %) would completely refrain from giving extra

marks due to economic hardships. Similarly, (40 %) of the teachers did not ethically accept bumping up students’ marks to make up for temporary circumstances. On the other hand, (40 %) were not sure of how to perceive such cases. As far as practices are concerned, only (46.7 %) made the decision to avoid giving an unprivileged student the unfair advantage of getting extra marks. Even though hiding students’ identity during grading students’ work to exclude any chances of bias or favouritism-was considered ethical by the majority of the teachers (66.7 %), only (26.7 %) admitted that they would regularly follow that practice in real classroom conditions.

5.3 Relationship between Beliefs/Practices and Other Factors

EFL teachers in this study possessed quite distinct ESL teaching and assessment experience and received different types of training. Teachers were divided according to their (a) years of teaching experience (1–9, 10–19 and 20–30); (b) years of assessment experience (1–10, 11–15 and 16–20) and according to the (c) type of assessment training they were exposed to (no training, pre-service, in-service and other training strategies). Differences between mean scores on both ethicality and practice were examined using 3×1 univariate ANOVA (F) tests to look at each dependent variable (ethicality and practices) to see if the three independent variables have a significant impact on them as displayed in Table 10.

Table 10 shows that except for the main effect of teachers’ experience on practice, no main effects were found for the three independent variables on both

Table 10 Univariate analysis of variance: main and interactional effects

Source	Dependent variable	Sum of squares	d.f.	Mean squares	F	P	Partial eta squared
Teaching exp1.	Ethicality	27.678	2	13.83	0.450	0.646	0.110
	Practice	309.875	2	154.9	8.502	0.003*	0.92
Assessment exp.	Ethicality	127.804	2	63.90	2.079	0.160	0.360
	Practice	29.687	2	14.84	0.815	0.461	0.163
Training	Ethicality	102.026	3	34.00	1.106	0.377	0.240
	Practice	127.074	3	42.35	2.324	0.116	0.472
Experience * exp_assess	Ethicality	135.458	2	67.72	2.203	0.145	0.379
	Practice	20.159	2	10.07	0.553	0.586	0.125
Experience * training	Ethicality	3.000	1	3.00	0.098	0.759	0.060
	Practice	21.333	1	21.33	1.17	0.296	0.173
Exp_assess * training	Ethicality	9.375	1	9.375	0.305	0.589	0.081
	Practice	0.453	1	0.453	0.025	0.877	0.053
Error		273.342	15	18.22			

Note *Interaction between two or more variables. *Exp* experience; *assess* assessment

Table 11 Mean difference between years of teaching and practice in Tukey post hoc test

	1–9 years	10–19 years	20–30 years
1–9 years		7.1*	9.3*
10–19 years	7.1*		2.3
20–30 years	9.3*	2.3	

Note *The locations of significant group differences in both Tukey HSD post hoc multiple comparisons

teachers' ethicality and practice. In other words, no statistically significant differences were found between teachers of distinct assessment experiences or differences between teachers exposed to various types of training in terms of ethicality, $p = 0.160$ and $p = 0.377$ in both cases respectively. Similarly, participants of different assessment and training experiences did not exhibit tangible differences in terms of their assessment practice, $p = 0.461$ and $p = 0.116$ for training and assessment experience respectively. Noticeably, no interaction at 0.05 between the study independent variables was found.

Therefore, as indicated in Table 10, it appears that teachers' previous experience has significant univariate main effect on teachers practice, $F(2,15) = 8.5$, $p = 0.003$, partial eta squared = 0.092. This means that participants in the three groups with different years of EFL teaching experience varied significantly in their mean scores on practices of ethical assessment. To examine the location of group differences, the statistical procedures of post hoc multiple comparisons were applied, as this study did not propose hypotheses about specific group differences. The Tukey HSD test was used.

In the post hoc multiple comparisons test of this study, since group sizes were unequal, harmonic mean sample size was used. In terms of the relationship between EFL teachers' years of English teaching and ethical assessment practice, the results showed that teachers from 1 to 9 years of experience and those with 10–19 differed significantly at $p < 0.05$; teachers with 10–19 years of experience performed better ($M = 111.4$) than those with 1–9 years of experience ($M = 99$) with respect to practice. Certainly also, teachers with 20–30 years of teaching experiences performed better ($M = 113$) than those with 1–9 years of experience ($M = 99$). Yet, notably, there were no statistically significant differences between teachers with 10–19 and those with 20–30 years of experience in terms of assessment practice, $M = 223$, (Table 11).

6 Discussion

The current study aimed at identifying the consistency between the teachers' ethical beliefs and their classroom assessment practices. Generally speaking, although teachers seemed somehow aware of what constitutes ethical versus unethical assessment practices, a discrepancy between their ethical perceptions and the course

of actions they chose to adopt could be detected. In other words, EFL teachers' notions of ethical assessment did not significantly bear upon their assessment practices. Principally, the teachers have to face the main dilemma of striking a balance between providing maximum support to individual learners and being honest to ensure fairness and support long term learning. In addition, it can be induced that intuition and discretion were given precedence when judging assessment fairness. Therefore, when asked to provide justifications for their answers, teachers were unable to apply ethicality standards and they appeared to be more governed by official considerations. The teachers reported also that external factors, such as time and curriculum constraints and mandated assessment policies, might affect assessment fairness or their adherence to ethical beliefs. Notwithstanding these remarks, in some cases teachers seemed to resort to ethical behaviour even though they could not perceive the underlying ethical motive of their actions. It seemed that teachers never used reflection to think of their assessment-oriented practices.

In particular, issues of confidentiality and transparency were well dealt with by most teachers and somehow borderlines were drawn between what should be public and what should be private. Nevertheless, some discrepancies between teachers' convictions and actions could be discerned. Most teachers agreed to the ethicality of stating how a task will be graded; yet they did not have the same attitude about sharing rubrics with students which might be ascribed to the teachers' belief that grading is an exclusive teacher's responsibility. Moreover, the teacher's image as a guardian might have caused many teachers to think that only points of strength should be discussed with a student; that is why many teachers did not realize the unethicality of hiding any information from the student. Noticeably, although teachers' answers reflected their unawareness of the unethicality of disclosing confidential information about students' academic achievement, in practice most of them showed a tendency to keep students' information somewhat confidential. This indicates that teachers were in many cases driven by their rational intuition or "practical wisdom" in Tierney's (2010) terms.

As far as *consistency* or conformity between assessment and curriculum objectives is concerned, a moderate correlation could be discerned between what teachers believed and how they tended to act. Generally, the respondents acknowledged the importance of assessing students on material they knew students had mastered which was reflected in their practice. However, many areas of discrepancy existed between what teachers believed and how they tended to act. For instance, even though teachers believed that students should be exposed to activities that enable them to anticipate the format of the assessment procedures, some teachers refrained from adopting such practices. Interestingly, most teachers had blurred vision regarding the consistency between the course objectives and methods of assessment adopted which was reflected also in their practices. Sometimes both teachers' perceptions and practices reflect their lack of awareness of what constitutes ethical assessment. For instance, when it comes to grading students, teachers seemed not well cognizant that the scores the students receive should tightly reflect their mastery of the skills stated by the objectives.

The present study also gave some indication that ethical dilemmas centring on *score pollution* made up the majority of incidents. These findings are consistent with the findings of Green et al. (2007). Issues that did not yield a great deal of conflict were those related to training students on certain exam questions and the unjustifiable deduction of scores to minimize guessing. Furthermore, teachers tended to ethically justify practices that reflect incorporation of students' non-academic performance such as effort, participation, improvements, laziness (...) etc. This indicates that teachers normally do not always ensure that the assessment tools and grading procedures employed actually reflect students' targeted competences. Thus, perhaps with clearer guidelines about what constitutes score pollution and why score pollution is unethical, these ethical dilemmas would not be so prevalent.

As for *AfL*, teachers were inclined to use multiple methods of assessment, yet a disproportionate heavy weight was allotted to testing as the best method for assessing students. Peer evaluation and self-evaluation were looked at with a lot of suspicion and integrating them in the classroom assessment plan was considered unethical by nearly most of the teachers. Many teachers thought that it is important to give students tasks that suit them and that not all students should be tested in the same way. Discrepancy between teachers' convictions and actions were obvious in that the high ethical value they accorded to practices such as slowing the teaching pace to adapt to students' needs and conferencing with students is not transferred to their everyday practice. In sum, teachers' practices in this respect contracted the concept of *AfL*.

Regarding *equity*, teachers' seemed to hold a clearer vision regarding ethicality, even if they were hesitant to apply what they perceived in their daily practices. In other words, the ethicality or otherwise of some practices, such as addressing students' interests, providing more than one format for a test, avoiding clues, seemed to be well settled and agreed upon by most teachers. However, in spite of teachers' apparent ethical approach, their practices fell short of reflecting their way of thinking. This might be due to fact that teachers are constrained by many factors that direct their practices. For instance, teachers perceived that students' special needs should be addressed; yet practically they found it difficult to address these needs. One interpretation is that teachers might have felt that any adaption to the assessment process should be the responsibility of other stakeholders, rather than the teacher himself. In some cases, teachers' ethical practices were driven by their rational intuition, even if they were inconsistent with their ethical convictions. For example, most teachers avoided providing clues to weak students, even though they could not ethically justify their sound practices.

The results of the current study give also some indication that teachers' perceptions of ethicality was not affected by their teaching or assessment experience. This can be attributed to the fact that most training programs focus on the practicalities of the assessment process and pay no heed to ethical issues underlying teachers' actions. Nonetheless, it was proved that teaching experience has a significant impact on teachers' ethical practices regardless of the training received. This contradicts what was suggested by previous research that ethical reasoning in

assessment does not develop on the job (Green et al., 2007). Yet, it seems that subsequent to fifteen years of experience, teachers tended to get accustomed to certain practices and that no remarkable change in their ethical decisions can be discerned.

7 Recommendations and Limitations

Results of the current study imply that many areas were considered controversial for most teachers. One of these areas was using multiple forms for assessing students. Another issue was consistency between the assessment methods used and the curriculum objectives and classroom activities. Equity issues also seem to be blurred for most teachers. Most teachers tended to adopt an over protective stance towards students regardless of whether or not this stood in sharp contrast to their own beliefs of what constitutes ethical assessment. Ensuring equity by avoiding bias toward certain groups such as students with limited ability or disability is also not well substantiated for teachers. In other words, although some rules were morally self-evident for the teachers, discrepancies between what teachers believed to be fair and what they got used to doing in class was obvious.

The generalizability of the specific results of this study may be limited by its use of a self-report survey and the limited number of the participating sample. Future studies may use multiple methods of data collection including classroom observation, analysis of teacher-made tests, and teacher interviews to validate teacher self-reports. In the future, also, the survey should be sent to a more representative sample selected from a variety of geographic regions across the country. The current data suggest that more time needs to be spent in confronting the ethical dilemmas of assessment and methods of approaching and resolving these dilemmas.

Results of the current study imply that general measurement training by itself cannot compensate for novices' lack of experience in terms of fair assessment. Nevertheless, the findings testify to the value of training that is particularly focused on fair assessment and ethicality dilemmas.

In light of previous results, it is recommended that teachers should be directed to put into consideration score pollution issues by providing clearer guidelines about what constitutes score pollution and why score pollution is unethical. Furthermore, explicit instruction in ethical concepts, such as equity, consistency, transparency and confidentiality, ought to be part of teacher pre-service training program as well in-service programs with ample chances for putting these ethical codes into practice by directly relating to the daily work in which teachers engage. Thus, the current study suggests pre-service and in-service training should address the issue of how to strike a balance between knowing a lot about students and avoiding biases. Teachers' awareness of the discrepancy between their roles as assistants to students and their roles as agents who takes part in establishing an accountable educational system should as well be raised.

Accordingly, continued research is needed to define more clearly the ethical issues teachers face as regards assessment. In addition, self-reflection practices should be encouraged among teachers; this can be accomplished by requiring teachers to report their regular ethical dilemmas pertinent to classroom assessment using reflection logs, diaries or group discussion.

Appendix

Teacher’s perception and practice survey

	How often you do that			Ethicality		
	U	S	N	E	SE	U
<i>A. Preparation for assessment</i>						
1. I state how I will grade a task when I assigns it (transparency+)						
2. I spend a class period to train students on test-taking skills (e.g., not spending too much time on one question, eliminating impossible answers, guessing) (consistency+)						
3. To prepare students to an upcoming test, I administer a parallel form of the test. The parallel form is another version of the test; however, the items are not the same as those on the final form of the achievement test (consistency+)						
4. Based on my review of a university final test framework, I create learning activities with specific questions that are included in the annual achievement test (score pollution)						
5. I don’t assess students until I make sure that I comprehensively covered the material and that students possess all the skills needed (assessment for learning)						
6. I tell students what materials are important to learn in preparing for a classroom assessment (score pollution)						
7. To minimize guessing, I announce that I will deduct more points for a wrong answer than for leaving the answer blank (pollution)						
<i>B. Assessment development</i>						
8. I use a previously designed test without referring to the objectives of the syllabus (consistency–)						
9. I assess my students’ knowledge by using many types of assessments: multiple-choice tests, essays, projects, portfolios (equity+)						
10. When I develop a rubric for correcting students’ written composition I hide the information about it as highly confidential (transparency–)						

(continued)

(continued)

	How often you do that			Ethicality		
	U	S	N	E	SE	U
11. In a reading test, I include texts that are relevant to the interests of various students' sub-groups (equity+)						
12. In a vocabulary test, I use assessment methods that students have never encountered before, for examples, drawing, giving examples, fill in a table (consistency-)						
13. I make sure that the activities included in a test were quite similar to activities presented in class (consistency+)						
14. I assess oral proficiency only through observing students during classroom discussion (consistency-)						
15. For the final exam, I use a few surprise items about topics that were not on the study guide (consistency-)						
16. If I have a blind student in my class, I design a recorded version of the test (equity+)						
17. I use more than one format of the same test to prevent cheating (equity+)						
18. In a vocabulary test, I put some clues in each item to help students find the answers easily (score pollution)						
19. For MCQ, I try to make the correct answer longer than others to help weak students answer better (equity-)						
<i>C. Administering assessment</i>						
20. If I notice that a student has skipped a question. I stop at his/her desk and show the him/her where to record the answer he is working on (equity-)						
21. While administering a test, when I notice that a student has missed a problem that he obviously knows, I stand by the student's desk, taps my finger by the incorrect problem, shake my head, and walk on to the next desk (equity-)						
22. While applying a test, I remind any student who stumbles on a question of what we learned by giving him or her a hint (equity-)						
23. Upon students' request, I would translate a difficult word that hinder students' understanding of a reading comprehension text (equity-)						
24. On a final exam, I would read all the test instruction orally with some emphasis on the key parts to help all students answer the questions easily (score pollution)						
25. For a slow student, I allot extra test time even if the test time has passed (equity)						
26. In a listening comprehension test, I read the text loudly to all students trying to highlight key parts by using a higher voice tone (score pollution)						

(continued)

(continued)

	How often you do that			Ethicality		
	U	S	N	E	SE	U
27. I would allow a student with a learning disability, i.e., a blind student, to use a tape-recorder when the student answers the essay questions on a grammar test (equity+)						
<i>D. Grading and feedback</i>						
28. I lower grades for late work by one score or more for each day (score pollution)						
29. I consider student effort when determining grades class (score pollution)						
30. In case of teaching two or more classes, I try be less strict in grading a class whose students I believe are weaker or slower (equity-)						
31. In an advanced reading class, I would assess reading based on students' final semester grade on two multiple choice tests (consistency-)						
32. For a group project, I base each student's grade on the group's product and a heavily weighted individual component (score pollution)						
33. To encourage lively discussion in English, I count class participation as part of the final grade (multiple assessment)						
34. I weigh tests heavily in determining students' final grades compared with other methods, i.e., homework, discussion, projects, presentation (multiple assessment-)						
35. I would give a student an F for the course because he/she missed the final exam (score pollution)						
36. I use student's peer assessment as a part of a final grade on an oral report (multiple assessment+)						
37. I lower class grades for disruptive behavior (score pollution)						
38. I would give extra scores to a student to make up for his/her underprivileged economic conditions (equity-)						
39. If I know a student had a bad week because of problems at home, I would bump his/her participation grade up a few points to compensate for his bad score on a quiz (equity-)						
40. If I believe that that students' work is rarely perfect in one of classes, I would make the decision of assigning very few grades of "A" to my class (equity-)						
41. I would change one student's course grade from a B+ to an A because tests and papers showed he/she had mastered the course objectives even though he had not completed some of his homework assignments (score pollution)						

(continued)

(continued)

	How often you do that			Ethicality		
	U	S	N	E	SE	U
42. I would offer extra credit opportunities to all the classes I teach except the advanced class (equity-)						
43. I hide the identity of the students (by concealing the name) whose essay test I'm grading so I won't identify them (equity+)						
44. I use peer evaluation with reference to certain rubric to help correct writing essays quickly (assessment for learning)						
<i>E. Communication of results and feedback</i>						
45. To enhance self-esteem, I address only students' strengths when correcting students' writing (transparency-)						
46. I spend time conferencing with each student to explain points of strength and weakness in their writing performance (assessment for learning)						
47. Based on the students' results, I would slow down my teaching pace to adapt to students' needs (assessment for learning)						
48. To motivate students to perform better, I would announce that I'm passing out scored tests to students in order of points earned, from the top score to the bottom score (confidentiality-)						
49. To calm the fears of distraught parents, I would compare their child's achievement scores with the results of the student's cousin who is also in the class (confidential-)						
50. I categorize students by labelling them to low level, high level, at risk (equity-)						

Usually (U); sometimes (S); never (N); ethical (E); somewhat Ethical (SE); unethical (U)

References

Airasian, P. (2005). *Assessment in the classroom: A concise approach* (2nd ed.). Boston, MA: McGraw-Hill Company.

Blanchard, J. (2008). Learning awareness: Constructing formative assessment in the classroom, in the school and across schools. *Curriculum Journal*, 19(3), 137–150. doi:10.1080/09585170802357454

Brookhart, S. M. (2004b). Classroom assessment: Tensions and intersections in theory and practice. *Teachers College Record*, 106(3), 429–458. doi:10.1111/j.1467-9620.2004.00346.x

Bursuck, W. D., Munk, D., & Olson, M. (1999). The fairness of report card grading adaptations: What do students with and without learning disabilities think? *Remedial and Special Education*, 20(2), 84–92. doi:10.1177/074193259902000205

- Buzzelli, C., & Johnston, B. (2002). *The moral dimensions of teaching: Language, power, and culture in classroom interaction*. London: Routledge Falmer.
- Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.) (pp. 221–256). Westport: American Council on Education & Praeger.
- Chan, Y. (2008). Elementary school EFL teachers' beliefs and practices of multiple assessments. *Reflections on English Language Teaching*, 7(1), 37–62. Retrieved from: <http://www.nus.edu.sg/celc/research/reft.php>
- Chang, Ch-W. (2006). Teachers' beliefs towards oral language assessment in Taiwan collegiate EFL classrooms. A paper presented at 2006 International Conference on English Instruction and Assessment. Retrieved from: http://flcccu.ccu.edu.tw/conference/2005conference_2/download/C03.pdf
- Dann, R. (2002). *Promoting assessment as learning: Improving the learning process*. London: Routledge Falmer.
- Educational Testing Service. (2002). *ETS standards for quality and fairness*. Princeton: Educational Testing Service.
- Gebriel, A., & Brown, G. T. L. (2013). The effect of high-stakes examination system on teacher beliefs: Egyptian teachers' conceptions of assessment. *Assessment in Education: Principles, Policy and Practice*, 21(1), 16–33. doi:10.1080/0969594X.2013.831030
- Gipps, C., & Murphy, P. (1994). *A fair test? Assessment, achievement and equity*. Buckingham: Open University Press.
- Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through pre-service teacher education. *Teaching and Teacher Education*, 21(6), 607–621. doi:10.1016/j.tate.2005.05.001
- Green, S., Johnson, R., Kim, D., & Pope, N. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education*, 23, 999–1011. doi:10.1016/j.tate.2006.04.042
- Hidri, S. (2015). Conceptions of assessment: Investigating what assessment means to secondary and university teachers. *Arab Journal of Applied Linguistics*, 1(1), 19–34.
- International Language Testing Association. (2000). *Code of ethics for ILTA*. Retrieved on March 30, 2009 from <http://www.iltaonline.com/code.pdf>
- JCSEE. (2003). *The student evaluation standards*. Arlen Gullickson, Chair. Thousand Oaks, CA: Corwin.
- Lu, Ai-ying. (2003). *Teachers' beliefs and classroom assessments: A case study of two university instructors of English* (Unpublished master's thesis). National Taiwan Normal University, Taiwan, R.O.C.
- McMillan, J. (2007). *Formative classroom assessment: Theory into practice*. New York: Teachers College Press.
- Messick, S. (1995). Validity of Psychological Assessment: Validation of Inferences from Persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. Retrieved from: <http://dx.doi.org/10.1037//0003-066X.50.9.741>
- Payne, D. A. (2003). *Applied educational assessment* (2nd ed.). Belmont, CA: Wadsworth.
- Pelly, C., P., & Allison, D. (2000). Investigating the views of teachers on assessment of English language learning in the Singapore education system. *Hong Kong Journal of Applied Linguistics*, 5(1), 81–106.
- Plake, B., & Jones, P. (2002). *Ensuring fair testing practices. The responsibilities of test sponsors, test developers, test administrators, and test takers in ensuring fair testing practices*. A paper presented at the February 2002 meeting of the Association of Test Publishers, Carlsbad, CA. Retrieved from: <http://www.testpublishers.org/assets/documents/Ensuring%20Fair%20Testing%20volume%204%20issue%201%20070202.pdf>
- Pope, N., Green, S., Johnson, R., & Mitchell, M. (2009). Examining teacher ethical dilemmas in classroom assessment. *Teaching and Teacher Education* 25, 778–782. doi:10.1016/j.tate.2008.11.013
- Popham, W. J. (2000). *Modern educational measurement: Practical guidelines for educational leaders*. Needham, MA: Allyn & Bacon.

- Shepard, L. (2005). Linking formative assessment to scaffolding. *Educational Leadership*, 63(3), 66–70.
- Shepard, L. (2007). Formative assessment: Caveat emptor. In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 279–303). Mahwah, NJ: Lawrence Erlbaum Associates.
- Simon, M., Chitpin, S., & Yahya, R. (2010). Pre-service teachers' thinking about student assessment issues. *International Journal of Education* 2(2), 1–20. doi:10.5296/ije.v2i2.490
- Stiggins, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758–765. doi:10.1177/003172170208301010
- Szpyrka, D. (2001). Exploration of instruction, assessment, and equity in the middle school science classroom. *Dissertation Abstracts International*, 62(10), Section: A, 3287–3548.
- Tierney, R. (2010). *Insights into fairness in classroom assessment: Experienced English teachers share their practical wisdom* (PhD dissertation). University of Ottawa, Canada (UMI No 69109).
- Volante, L. (2006). Reducing bias in classroom assessment and evaluation. *Orbit*, 36(2), 34–36. Retrieved from: <https://ezpa.library.ualberta.ca/ezpAuthen.cgi?url=search.proquest.com/docview/213742716?accountid=144744>
- Yip, D. & Cheung, D. (2005). Teachers' concerns on school-based assessment of practical work. *Journal of Biological Education*, 39(4), 156–162. doi:10.1080/00219266.2005.9655989
- Zhang, Z., & Burry-Stock, J. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16(4), 323–342. doi:10.1207/S15324818AME1604_4
- Zoeckler, L. (2005). Moral dimensions of grading in high school English. *ProQuest digital dissertations*. UMI No. AAT3183500.