

Ethical assessment practices in English classes: A focus on experience and gender

Hadi Azimi¹

*Assistant Professor, Shahid Beheshti University of Medical
Sciences, Tehran, Iran*

Received on March 28, 2012

Accepted on September 8, 2012

Abstract

Assessment, as a key component of education, has long been a matter of concern to teachers and it receives even more significance when perceived from the view point of ethics. The combination of the two concepts, i.e. ethical assessment, is difficult to define and complicated to measure. But prior to all these comes how teachers *perceive* it. The present research reports on a study of English teachers' perception of ethical assessment. Following Green, Johnson, Kim, and Pope (2007), 108 English teachers were asked to state their opinions about ethicality or unethicity of 40 prevalent assessment practices classified into seven themes: *test preparation, communication about grading, multiple assessment opportunities, test administration, grading practices, confidentiality, and neutrality*. The findings suggest that respondents displayed consensus on only two fifth of the scenarios. To delve into the reasoning behind participants' choices, 2 participants were interviewed for each scenario and the findings, also, were thematically compared with those reported in literature.

Keywords: ethics, assessment, grading practice, English teachers

¹ Corresponding author: hadiazimi@sbm.ac.ir

1. Introduction

A global definition of assessment reads “the process of gathering information to make informed decisions.” (Anderson, 2003, p. xi) Along with this process resorting to different methods and procedures is inevitable, since different students hold various proclivities (Eisner, 1994). Meanwhile, proper assessment is highly appreciated because it provides useful insights about the success of the program and offers feedback to the teachers (Antman, 2007; Asmus, 1999; Eisner, 1994) and plays the role of gate-keeping as well as educational temperature-taking (Eisner, 1994). The significance of assessment and its influences in the educational milieu can be summarized in Gipps’ (1995) beginning lines of his chapter on ‘ethics and equity’. He stated that assessment:

...is a powerful tool: it can shape curriculum, teaching and learning; it can affect how pupils come to see themselves both as learners and in a more general sense as competent or not; through labeling and sorting pupils (certificating and selecting) it affects how pupils are viewed by others; it controls access to further education and high status career. (p. 144)

Bearing such influences in mind, researchers have frequently appreciated the sound validity and reliability of assessment practices (Cheng, Rogers, & Hu, 2007; Haladyna, Nolen, & Hass, 1991; Gipps, 1995; Mehrens & Kaminski, 1989; Messick, 1989; Boyle & Radocy, 1987; Wiggins, 1989). Now, soundness of these criteria is perceived from different perspectives including ethics and fairness.

Ethics is simply defined as “what is good and bad and with moral duty and obligation” (*Webster’s Ninth New Collegiate Dictionary*, 1988, p. 426). Ethics is different from morals in that the former has to do with a professional context, so it is an extension of morals (Melo, 2003). To be more precise, the question of ethicality or unethicity is answered by viewing a behavior to be along with or opposed to one’s obligations and choice (Green, Johnson, Kim, & Pope, 2007). Discussions of ethics have been with assessment for about 40 years now (Cole & Zieky, 2001) after it was educationally

revealed that practitioners were susceptible to unethical practices leading to misjudgments and unjust decision makings (Schmeiser, 1995). It is claimed that a major part of these unethical practices stem from the existence of conflicting norms and complex situations with regard to teaching (Colnerud, 1997; Schmeiser, 1995). In this regard, ethical principles can help improve ethical assessments (Green et al, 2007; Plake & Impara, 1997; Stiggins, 1999).

One of the good representations of classroom practices, which has recently attracted researchers' attention (Green et al., 2007) is grading practices. Like other aspects of teaching and assessment, grading is highly susceptible to unethical practices "because at the core of the teaching task is the relationship between teacher and student. This relationship is of an unequal and dependent nature" (Melo, 2003, p. 179). Scheire (2008), among others, called for training the staff for proper ethical grading procedures. Also, Moore (1993) stated that "classroom educators are not prepared to implement appropriate and acceptable test preparation and test administration practices" (p. 1), that is to say, without proper awareness and education of the ethical considerations, ethics of the assessment will be under serious doubts. There are many codes for the practice of ethics in assessment, especially in US (Green, et al., 2007); however, because values and ethical considerations may be defined differently in different contexts, investigation of various dimensions of these principles seems quite indispensable in various settings, including the present case of Iranian context. Yet, there is no officially stated set of codes or principles with regard to ethical assessment to follow in Iran. For that matter, the present study looks for consensus among teachers in terms of their judgment of some sample grading practices as a primary step to study ethical practices.

2. Literature Review

The practice of ethics is widely appreciated in assessment literature (Baumgart, 1996; Gipps, 1994; Popham, 2000). Some research addressed the fundamental notion of teachers' reflections in ethical dilemmas (Shapira-Lishchinsky, 2011) as the way teachers reflect

in/on such situations is informative of their other critical actions, such as assessment. Schmeiser (1995) stated that it is students' right "to be instructed with effective and fair methods by professionals who use appropriate grade level materials and activities that fit the learning context." (p. 318) He referred to various associations in US that developed a position paper on fair and wise use of tests. Also, Stiggins (1999) pleaded for excellence in classroom assessment because "the academic well-being of the student hangs in the balance" (p. 27). In fact, topics such as grading and standardized testing are among the frequent points of conflict in assessment that generate critical incidents (Pope, Green, Johnson, & Mitchell, 2009). Also, Pantic and Wubbles (2012) explored the possibility whether teachers' beliefs and moral values could affect the student-teacher relationship, and thereafter, the teaching and assessment practices; the findings supported existence of a causal relationship. Genesee and Upshur (1996), dealing with classroom-based assessment, briefly discussed the ethical concerns teachers should follow while checking students' journals. They are advised to provide sympathetic and supportive responses, if necessary. At the same time, teacher's practice is advised to be informed by ethical theories to be able to successfully solve their everyday encounter with ethical and moral dilemmas (Malloy & Hansen, 1995). On the whole, presently, literature is viewed alert toward awareness in assessment and teaching (Campbell, 1993; Colnerud, 1997; Jackson, Boosrom, & Hansen, 1993; Melo, 1993).

2.1 Ethical Guidelines in Assessment

In addition to the benefit of providing a sense of identity to the profession, which "invariably includes the mention of an adherence to a code of ethics" (Melo, 2003, p. 180), in order for assessment to be ethical, curriculum developers need to make sure practitioners are well-aware of the ethical considerations and, in the next level, to ensure this ethical knowledge is properly applied. One big step is setting guidelines to follow. These guidelines, as Green et al. (2007, p. 999) cite the Joint Committee on Standards for Educational Evaluation (JCSEE, 2003), are offered as "principles that should guide and govern student evaluations (JCSEE, p. xx)." As Baumgart

(1996) stated, these codes offer teachers great instructions for more ethical practice. Furthermore, these standards are provided so that the “evaluation be ethical, fair, useful, feasible, and accurate” (JCSEE, p. 3, cited in Green et al., 2007). Green et al. (2007) introduced many such guidelines suggested in the literature (Haladyna et al., 1991; Mehrens & Kaminski, 1989; Popham, 1991; Shortt, Hallett, Spendlove, Hardy, & Barton, 2012) confessing these guidelines are “general and almost superficial” (Gipps, p. 144) and thus need further explanations and logical interpretations by teachers. Later, they stated that “ethical standards related to testing are not consistently defined or agreed upon (Thorndike, Cunningham, Thorndike, & Hagen, 1991; and see Kilian, 1992, in response to Popham, 1991)” (p. 1000).

Ethical guidelines stated in the literature are twofold: some are dictations of certain associations or departments affiliated to educational policies in a state or a country, such as the *Principles of Fair Student Assessment Practices for Education in Canada* by the Joint Advisory Committee (1993). An example of the proposed guidelines is the 3rd principle under Collection Assessment Information which reads:

3. In assessments involving observations, checklists, or rating scales, the number of characteristics to be assessed at one time should be small enough and concretely described so that the observations can be made accurately.

Student behaviors often change so rapidly that it may not be possible simultaneously to observe and record all the behavior components. In such instances, the number of components to be observed should be reduced and the components should be described as concretely as possible. One way to manage an observation is to divide the behavior into a series of components and assess each component in sequence. By limiting the number of components assessed at one time, the data and information become more focused, and time is not spent observing later behavior until prerequisite behaviors are achieved. (p. 7)

Other codes are speculations of researchers or experts in testing and ethics fields published in the journals, like the ones Stiggins (1999)

proposed as seven assessment competencies teachers need to possess for a proper assessment (Table 1).

Table 1: Stiggins' (1999) seven assessment competencies

Seven Assessment Competencies	
1.	Connecting assessment to clear purposes
2.	Clarifying achievement expectations
3.	Applying proper assessment methods
4.	Developing quality assessment exercises and scoring criteria and sampling appropriately
5.	Avoiding bias in assessment
6.	Communicating effectively about student achievement
7.	Using assessment as an instructional intervention

Cole and Zieky (2001), too, claimed that “in spite of more than 30 years of efforts, there is still no generally accepted definition of fairness with respect to testing and no measure that can prove or disprove the fairness of a test.” (p. 369) As a result, they proposed 4 criteria to call a test fair. To them, a fair test is the one which: 1) ameliorates group differences, 2) provides opportunities to perform, 3) deters misuse, and 4) accommodates individual differences (p. 376). But still the broadness of the guidelines is a big flaw.

Green et al. (2007), after reviewing a host of texts, ranging from professional journals to assessment texts, identified “two general guiding principles to capture the essence of the ethical concerns in these documents.” (p. 1001): (1) Do Not Harm, and (2) Avoid Score Pollution. The first general principle, based on Taylor and Nolen (2005), is prevalent in many other professions like medicine (Jonsen, Siegler, & Winslade, 1998; Munson, 2000, both cited in Green et al., 2007), as well. What counts first is defining harm:

or choosing between different harms in the relationship between teacher and student. This principle stems from the basic premise that ethical guidelines must protect the rights of individuals affected by an evaluation. (Green et al., 2007)

They adapted the second general principle from suggested guidelines by both Popham (1991) and Haladyna, et al. (1991), which has to do with “ethical standardized test preparation”. “Both authors suggest that any practice that improves test performance without concurrently increasing actual mastery of the content tested produces score pollution.” (Green, et al., p. 1001) These guidelines made up the backbone for their paper, and for the present study as well.

Bachman and Palmer (1996) defined the scoring methods as “the criteria by which test takers’ responses are evaluated and the procedures followed to arrive at a score.” (p. 193) They classified the measurement processes to arrive at certain scores into three steps: defining the construct theoretically, defining the construct operationally, and the measurement process, respectively. The point which relates this discussion to our purpose is that in each of these levels unethical practices may pervade, as Moore (1993) claimed that “classroom educators are not prepared to implement appropriate and acceptable test preparation and test administration practices.” (p. 1) But first of all, one should examine the extent to which teachers apply these guidelines in their classroom assessment practice on a consistent basis. Green, et al. (2007), as a model study in this area, reported that agreement among the participating teachers in terms of the ethicality or unethicality of the interrogated grading practices was found only in fewer than half of the presented scenarios, revealing lack of consensus in the profession. The present study, which is a replication and extension of the study carried out by Green et al. (2007), is an exploration of the level of agreement or disagreement among English teachers concerning their judgments of the ethicality or unethicality of some common grading scenarios prevalent in English classes.

3. Purpose of the Study

This study was carried out in order to mainly understand how English teachers in Iran perceive different assessment practices ethical or unethical. This is the first step in a chain of studies to look into teachers’ practice of ethical assessment. In other words, understanding the extent to which teachers see ethical assessment

ethical and, for that matter, unethical assessment unethical, provides valuable insights on how further measures should be taken. Moreover, the hypothesis that gender and experience may affect the way teachers perceive ethical assessment is appealing enough. Insights on how these variables can be related to teachers' perception of ethical assessment can promise further valuable research and pedagogical implications.

As a result, the present study was carried out in order to answer the following research questions:

1. Do novice and experienced English teachers perceive ethical practices similarly?
2. Do male and female English teachers perceive ethical practices similarly?

If the answers to these two questions are negative, the following question will be inevitable:

3. Why does English teachers' perception of ethical practices differ?

4. Method

4.1 Participants

By making use of 'snowball technique' of participant sampling (Milroy & Gordon, 2003, p. 32), 108 English teachers were selected to compose the participant pool for this study. All participants were holding MA or PhD in English related majors, i.e. Applied Linguistics, General Linguistics, and English Literature at the time of data collection. The participants were of both genders (38 male and 70 female participants) and with a range of teaching experience from 1 year to 27 years. The participants were then divided into two major groups in terms of their teaching experience, believing this variable can have strong influence on their grading practices. Based on Mackey, Polio, and McDonough (2004), Avalos and Aylwin (2007), Watzke (2007), and Scherff (2008) decision was made to

Ethical assessment practices in English classes

choose the 3rd year of teaching experience as the borderline between novice and experienced teachers (Table 2). Gender was also controlled to further delve into the elicited data.

Table 2: The frequency of participants (gender and experience)

		Frequency	Percent
Gender	Male	38	35.1
	Female	70	64.9
Experience	Novice	72	66.7
	Experienced	36	33.3

4.2 Instrument

In order to check the participants' judgments on ethicality or unethicity of various grading practices, following Green et al. (2007), more than 49 scenarios on grading practices were primarily developed and/or adapted from Green et al. (2007) paper. These scenarios were developed based on resources from literature on ethical assessment and different codes of ethics like *The Student Evaluation Standards* (JCSEE, 2003) and *The Principles for Fair Student Assessment Practices for Education in Canada* (Joint Advisory Committee, 1993) as well as personal experiences of the authors. Most of the items adapted from Green et al. (2007) were modified and rewritten to suit the context of the study and some of the redundant items were omitted. The final scenarios briefly described 46 situations in which a teacher was engaged in the process of assessing a student. The assessment practice whatsoever might or might not be ethical to different participants. For instance, item no. 23 reads "For the final exam of a Reading Comprehension course, a teacher always uses a few unseen items about topics that were not on the syllabus." Participants may choose either of the Ethical or Unethical options.

4.3 Data Collection Procedure

After the 46-scenario checklist was prepared, 4 experts in Applied Linguistics and Education (including the main author in Green, et

al. paper) were asked to review the scenarios. After this phase, 4 scenarios were omitted due to redundancy or irrelevance to the context. Yet again the modified checklist was administered to a group of 28 participants with identical characteristics and, as a result, one item was omitted and some of others were modified. The final 40-item checklist was administered to the target participants to check under either *ethical* or *unethical* columns for each scenario, and the elicited information was submitted to the SPSS for further statistical analysis.

4.4 Data Analysis

Based on the relevance of different items and to be as close to the classification reported in Green et al. (2007) as possible, all 40 scenarios were classified into 7 categories encompassing all the procedures normally implemented from preparation to reporting the results of a test: *test preparation*, *test administration*, *multiple assessment opportunities*, *communication about grading*, *grading practices*, *neutrality*, *confidentiality*. The thematic classification of the scenarios into these categories is identical to the one reported in Green, et al. (2007) helping us to be able to cross compare the results.

The criterion for determining 'high agreement' among participants was decided to be 80 percent of the answers. In other words, in case at least 80% of respondents stated a scenario was ethical/unethical, that item displays significant agreement among the participants in the group under the study. For instance, 90.7% of the respondents stated that it was ethical when a teacher always used a few unseen items about topics that were not on the syllabus while giving the final exam of a Reading Comprehension course (Item 23), and only 9.3% of the participants believed this was an unethical practice. To the same token, high agreement was significantly observed among participants when they were asked to judge on a teacher's practice who compared a child's achievement scores with the results of the student's cousin who was also in the class in order to calm the fears of the worried parents (Item 5). 81.5% of the respondents checked teacher's practice in this scenario as an unethical one.

To see on which bases participants disagree, following Green, et al. (2007), those items with percentages from 50% to 70% were categorized as ‘disagreement’. As a result, participants disagreed on the practice of a teacher who based students’ final semester grade on 2 multiple choice tests (Item 27). In this case, 47.2% of all respondents stated this practice was ethical and 52.8% viewed it as an ethical decision.

5. Results and Discussion

5.1 Analysis within Content Categories

In the first category, *test preparation*, the level of agreement ranged from 53.7% to 83.3% (Table 3). Out of 5 scenarios in this category, only one displayed agreement among the participants. 83.3% of respondents believed it was ethical for a teacher to spend a class period training his students in test-taking activities (Item 21). Other items displayed disagreement among the participants, i.e. administering a parallel test (Item 2), adding TOEFL vocabulary to classroom test (Item 9), creating learning activities with specific exam questions (Item 13), and proving students with 150 questions to read for the final exam (Item 28). The existing states of agreement and disagreement were also true for the two levels of experience and gender variables.

Table 3: Percentage of teachers indicating the ethicality of assessment practices in Test Preparation

Item	Scenarios about Test Preparation	Respondents’ answers	Gender		Experience	
			Male	Female	Novice	Experienced
2	A 3 rd grade senior high school English teacher administers a parallel form of a standardized test to her students in preparation for the state testing which is alike across the country. The parallel form is another version of the state test that assesses the same content; however, with different items.	Ethical	55.2	52.8	51.3	58.3
		Unethical	44.8	47.2	48.7	41.7

Azimi

9	A teacher adds vocabulary words from TOEFL to classroom vocabulary tests.	Ethical	34.2	34.2	33.3	36.1
		Unethical	65.8	65.8	66.7	63.9
13	Based on his review of the previous national high school final exams, a senior high school teacher creates learning activities with specific exam questions that are usually present in the final exams.	Ethical	78.9	60	62.5	75
		Unethical	21.1	40	37.5	25
21	A teacher spends a class period to train his students in test-taking skills (e.g., not spending too much time on one problem, eliminating impossible answers, guessing).	Ethical	84.2	82.8	87.5	75
		Unethical	15.8	17.2	12.5	25
28	During the semester, a teacher gives 150 questions to the students 15 of which will be chosen for the final achievement exam, as she announces.	Ethical	57.8	61.4	58.3	63.8
		Unethical	42.2	38.6	41.7	36.2

In the second category, all the items revealed high agreement among respondents. The range of agreement percentages for *test administration* was from 80.5% to 90.7%, which shows highly coherent test administration perceptions in the universities although only 3 scenarios were provided for judgment (Table 4). In this category, reminding the students where and how to write the answers at the final exam (Item 4), tapping on the incorrect answer for reanalysis (Item 17), and letting the students bring an already written essay to the final exam session of an essay writing course as part of the final score (Item 25) were believed to be ethical. In this category, too, both novice and experienced and female and male teachers followed the pattern disclosed in the category.

Table 4: Percentage of teachers indicating the ethicality of assessment practices in Test Administration

Item	Scenarios about Test Administration	Respondent	Gender		Experience	
		s' answers	Male	Female	Novice	Experienced

Ethical assessment practices in English classes

4	While administering an achievement test, a teacher notices that a child has skipped a problem and is now recording all his answers out of sequence on the answer form. The teacher stops at the child's desk and shows the student where to record the answer he is working on and instructs him to put the answer to each question with the same number on the answer sheet..	Ethical	94.1	88.5	91.6	88.8
		Unethical	5.3	11.5	8.4	11.2
17	While administering the final exam, a teacher notices that a child has missed a problem that the student obviously knows. The teacher stands by the child's desk, taps her finger by the incorrect problem, shakes her head, and walks on to the next desk.	Ethical	92.1	87.1	87.5	91.6
		Unethical	7.9	12.9	12.5	58.4
25	An instructor believes 90-minute time limit is not enough for students to prove their writing skill on an Essay Writing exam, so she asks students to write an essay on a shared topic, bring it to the final exam session, and attach it to their answer sheet as the major fraction of their final score.	Ethical	73.6	84.2	77.7	86.1
		Unethical	26.4	15.8	22.3	13.7

Participants divided themselves into two groups by showing agreement on two items and disagreement on the other two in *multiple assessment opportunities* category. As for the scenarios in this category, 82.4% of respondents stated that taking 3 quizzes during the semester

to make sure students had mastered the covered material(s) without incorporating the results in students' final scores, (Item 6) and use of many different assessment procedures were viewed as ethical (Items 34), but disagreement was elicited on the scenarios when students' final semester grades were based on 2 multiple choice tests (Item 27) and when assessment was solely based on observation for a methodology course (Item 38). Once more, neither gender nor experience showed to be affecting participants' judgments (Table 5).

Table 5: Percentage of teachers indicating the ethicality of assessment practices in M. A. Opportunities

Item	Scenarios about M. A. Opportunities	Respondent s' answers	Gender		Experience	
			Male	Female	Novice	Experienced

6	A Reading I instructor gives 3 quizzes during a semester to make sure students master the covered material(s), but she doesn't incorporate the results in students' final scores.	Ethical	78.9	84.2	81.9	83.3
		Unethical	21.1	15.8	18.1	16.7
27	A senior high school English teacher bases students' final semester grade on 2 multiple choice tests.	Ethical	42.1	50	48.6	44.4
		Unethical	57.9	50	51.4	55.6
34	A teacher assesses student knowledge by using many types of assessments such as multiple-choice tests, essays, projects, and portfolios.	Ethical	78.9	84.2	80.5	86.1
		Unethical	81.1	15.8	19.5	13.9
38	A Methodology instructor uses observations as the sole method to assess what students have learned.	Ethical	36.8	38.5	31.96	50
		Unethical	63.2	61.5	68.1	50

The scenarios in the category *communication about grading* challenged respondents' judgments about the ethicality of teachers' talks with students about the way grading was to be determined. Respondents showed agreement in 3 out of 4 scenarios (Table 6). The only item holding disagreement was item 12 in which a teacher tells students what materials are important to learn in preparing for a class test. Agreement for this category ranged from 55.6% to 90.7% when a teacher stated how she would grade a task when she assigned it (Item 10), a teacher's use of a few unseen items about topics that had not been stated on the syllabus (Item 23), and a teacher's giving students a 'criteria chart' that explained how final grades would be calculated (Item 31). Gender and experience did not affect the judgments either.

Table 6: Percentage of teachers indicating the ethicality of assessment practices in Grading Communication

Item	Scenarios about Grading Communication	Respondent s' answers	Gender		Experience	
			Male	Female	Novice	Experienced
10	A teacher states how she will grade a task when she assigns it..	Ethical	81.5	84.2	79.1	91.6
		Unethical	18.5	15.8	20.9	8.4
12	A teacher tells students what materials are important to learn in preparing for a class test.	Ethical	52.6	40	34.7	63.8
		Unethical	47.4	60	65.3	36.2
23	For the final exam of a Reading	Ethical	94.7	88.5	91.6	88.8

Ethical assessment practices in English classes

	Comprehension course, a teacher always uses a few unseen items about topics that were not on the syllabus.	Unethical	5.3	11.5	8.4	11.2
31	A teacher gives students a 'criteria chart' that explains how final grades are calculated.	Ethical	92.1	88.5	87.5	94.4
		Unethical	7.9	11.5	12.5	5.6

Grading practice is the most crowded category with seventeen scenarios. Agreement in this category ranged from 52.7% to 90.7% (Table 7). The present category, to our surprise, displayed poor agreement among respondents. In other words, teachers agreed on only 5 (out of 17) scenarios and the remaining 12 items proved teachers were split on their attitudes toward the ethicality of the assessment practices in terms of grading practices. The scenarios displaying agreement were: basing students' grades on the groups' product (Item 1), considering students' effort when determining grades (Item 16), a teacher's use of student peer ratings as 30% of the grade (Item 19), considering a student's growth in assigning grades (Item 22), and a teacher's heavily weighing term projects (Item 24). Experience and gender were consistent in most of the cases. Only female teachers and novice teachers in item 1, females in item 16, experienced teachers in item 19, and male teachers in item 24 revealed small insignificant fluctuations.

Table 7: Percentage of teachers indicating the ethicality of assessment practices in Grading Practice

Item	Scenarios about Test Practice	Respondent s' answers	Gender		Experience	
			Male	Female	Novice	Experienced
1	For a group project, a teacher bases each student's grade on the group's product plus evaluating group members individually.	Ethical	13.1	22.8	20.8	16.6
		Unethical	86.9	77.2	79.2	83.4
3	As a teacher finalizes grades, she changes one student's final grade from 16 to 17 because tests and papers showed the student had mastered the course objectives even though he had not completed some of his homework assignments.	Ethical	50	32.8	40.2	36.1
		Unethical	50	67.2	59.8	63.9
11	A teacher lowers grades for late work by one grade for each day. The due day was previously announced.	Ethical	55.2	54.2	52.7	58.3
		Unethical	44.8	45.8	47.3	41.7

Azimi

15	To encourage lively discussion in conversation classes, a teacher counts class participation as 30% of the final grade, i.e. 6.5 out of 20.	Ethical	47.3	55.7	48.6	61.1
		Unethical	52.7	44.3	51.4	38.9
16	A teacher considers student effort when determining grades.	Ethical	86.8	77.1	80.5	91.6
		Unethical	13.2	22.9	19.5	8.4
19	A teacher uses student peer ratings as 30% of the grade on the oral section of a Conversation course.	Ethical	84.2	82.8	88.8	72.2
		Unethical	15.8	17.2	11.2	27.8
22	A teacher considers a student's growth in assigning grades.	Ethical	86.8	92.8	94.4	83.3
		Unethical	13.2	7.2	5.6	16.7
24	A teacher weights term project heavily in determining students' grades.	Ethical	78.9	81.4	80.5	80.5
		Unethical	21.1	18.6	19.5	19.5
26	An instructor gives his BA students' essay type exam papers to his MA students in ELT for correction and grading.	Ethical	65.7	67.1	68	63.8
		Unethical	34.3	32.9	32	36.2
30	A teacher lowers grades for disruptive behavior.	Ethical	65.7	62.8	63.8	63.8
		Unethical	34.3	37.2	36.2	36.2
32	An instructor looks for only the key words in students' essay type answers. A student who has not used the key words will lose the point.	Ethical	52.6	61.4	52.7	69.4
		Unethical	47.4	38.6	47.3	30.6
33	Out of 20, a Phonology teacher specifies only 1 or 2 points to class participation and class attendance, and weights final exam heavily, i.e. 18 points, in assigning grades.	Ethical	57.8	64.2	62.5	61.1
		Unethical	42.2	35.8	37.5	38.9
35	A junior high school English teacher gives a student a zero as a homework grade for not returning a form requiring a parent's signature.	Ethical	42.1	47.1	43	50
		Unethical	57.9	52.9	57	50
36	A teacher keeps giving similar grades to a student of hers in different courses mainly based on her primary assessment in the very first course.	Ethical	65.7	52.8	52.7	66.6
		Unethical	34.3	47.2	47.3	33.4
37	A Study Skills instructor gives high grade to a student mostly because he has proved to be a fluent English speaker during the course.	Ethical	31.5	51.4	41.6	50
		Unethical	68.5	48.6	58.4	50
39	To minimize guessing, a Reading Comprehension teacher announces she will deduct more points for a wrong answer than for leaving the answer blank on the final exam.	Ethical	47.3	45.7	44.4	50
		Unethical	52.7	54.3	55.6	50
40	An Advanced Writing instructor divides the score on a paragraph-writing item into detailed elements constructing the text. So, a student will lose a point (out	Ethical	52	38.5	41.6	27.7
		Unethical	48	61.5	58.4	72.3

Ethical assessment practices in English classes

of 10) because he has not indented the first line, though the paragraph is sound and comprehensive.

In the next category, *neutrality*, showed a small range of agreement from 54.7% to 62.9% (Table 8). This category was the only category in which none of the scenarios was agreed upon. In fact, in the most agreed scenarios only 62.9% of the participants said they thought it would be ethical to add a few points to a student's final score when she came to know that child had a bad week (Item 29), which shows no classification of all the participants, neither male/female nor novice/experienced classifications, had consensus in terms of their beliefs towards ethicality of being neutral in such assessment practices. Other scenarios include a teacher's gender-biased grading (Item 14), a teacher's grading with knowledge of the students' identities on an essay type exam (Item 18), and a teacher's belief that students' works are rarely perfect to be graded 20 (Item 20). The highest percentage for agreement among subcategories was 68.5% for female teachers on item 29.

Table 8: Percentage of teachers indicating the ethicality of assessment practices in Neutrality

Item	Scenarios about neutrality	Respondent s' answers	Gender		Experience	
			Male	Female	Novice	Experienced
14	A female instructor gives higher scores to her female students irrespective of the male students' performance.	Ethical	47.3	37.1	36.1	50
		Unethical	52.7	62.9	63.9	50
18	A teacher always knows the identity of the student whose essay type test she is grading.	Ethical	50	42.8	40.2	55.5
		Unethical	50	57.2	59.8	44.5
20	Because of her belief that students' work is rarely perfect, a Conversation instructor rarely gives 20 as the final score.	Ethical	65.7	54.2	58.3	58.3
		Unethical	34.3	45.8	41.7	41.7
29	A teacher who knows a student had a bad week because of problems at home adds a few points to the student's participation grade to compensate for	Ethical	52.6	65.8	65.2	58.3
		Unethical	47.4	31.5	34.8	41.7

his bad score on a quiz.

The only item in *confidentiality* category which fell between agreement and split borderlines was item 8 (Table 9). 71% of the male participants and 72.2% of the experienced participants stated that passing out scored tests to students in order of points earned from the top score to the bottom score so as to motivate students was ethical.

Table 9: Percentage of teachers indicating the ethicality of assessment practices in Confidentiality

Item	Scenarios about Confidentiality	Respondent s' answers	Gender		Experience	
			Male	Female	Novice	Experienced
5	To calm the fears of worried parents, a teacher compares their child's achievement scores with the results of the student's cousin who is also in the class.	Ethical	18.4	18.5	19.4	16.6
		Unethical	81.6	81.5	80.6	83.4
7	A teacher discloses to the parents of a student their child's score on an intelligence test.	Ethical	89.4	84.2	90.2	77.7
		Unethical	10.6	15.8	9.8	22.3
8	To motivate students to perform better, a teacher always announces that she is passing out scored tests to students in order of points earned, from the top score to the bottom score.	Ethical	71	70	69.4	72.2
		Unethical	29	30	30.6	27.8

81.5% of respondents believed that comparing a student's scores with those of his cousin to calm the worried parents (Item 5) was unethical and 86.1% of them viewed a teacher's disclosing to the parents of a student their child's score on an intelligence test as ethical. Neither gender nor experienced affected the group judgment. Figure 1 displays and compares the percentages given for ethical and unethical assessment practice across 7 categories. As it is shown, the most references to ethicality and unethicality relate to *test administration* and *neutrality* themes, respectively.

5.2 Comparison between Novice and Experienced Teachers

In order to see if different teaching experience affects teachers' judgments' of ethicality of the scenarios, a '2 independent samples tests' was run to locate the likely difference between novice and experienced teachers. The results revealed that, statistically speaking, the two groups' stated ideas were identical in all but two items, i.e. items 12 and 19 where the two groups showed significant difference ($0.05 > p$) (Table 10). Consequently, teaching experience was not a determining factor in respondents' judgment of the ethicality/unethicality of assessment practices.

Table 10: Test statistics for 2 independent samples test for teachers' experience^a

	Scenario 12	Scenario 19
Mann-Whitney U	918.000	1080.000
Wilcoxon W	1584.000	3708.000
Z	-2.862	-2.181
Asymp. Sig. (2-tailed)	.004	.029

^a Grouping Variable: Experience

5.3 Comparison between Male and Female Teachers

Likewise, the possible effect of gender was subjected to the SPSS. In this analysis too, generally speaking, neither male nor female respondents showed significant difference across the 40 scenarios. Only in two items significant statistical differences were observed: in item 13, creating learning activities with specific exam questions that are usually present in the final exams, and item 37, giving high grade mostly because of language fluency for a Study Skills course (Table 11).

Table 11: Test statistics for 2 independent samples test for teachers' gender^a

	Scenario 13	Scenario 37
--	-------------	-------------

Mann-Whitney U	1078.000	1066.000
Wilcoxon W	1819.000	3551.000
Z	-1.985	-1.973
Asymp. Sig. (2-tailed)	.047	.048

^a Grouping Variable: Gender

5.4 Analysis across Content Categories

Table 12 summarizes the analysis of high agreement and high disagreement across seven categories. Based on the number of scenarios on which respondents had high agreement *test administration* category ranked first. All the three scenarios were highly accepted by the respondents (100%). The following categories were *communication about grading* (75%), *confidentiality* (67%), *multiple assessment opportunities* (50%), *grading practices* (29%), *test preparation* (20%), and finally *neutrality* with no agreed item, respectively from the highest to the lowest percentage of agreement. Items with high disagreement were also ranked, which revealed *neutrality* category with 100% of the items displaying high disagreement (4 items) and *test administration* and *confidentiality* categories with no scenario displaying high disagreement held respectively the most and the least items about which respondents split.

Table 12: Categories ranked by percent of items showing high agreement (80%+) and high disagreement (50-70%)

Category	Items showing high agreement	Items showing high disagreement
Test administration	3 (100%)	0
Communication about grading	3 (75%)	1 (25%)
Confidentiality ^a	2 (67%)	0
Multiple assessment Opportunities	2 (50%)	2 (50%)
Grading practices	5 (29%)	12 (71%)
Test preparation	1 (20%)	4 (80%)
Neutrality	0	4 (100%)

Ethical assessment practices in English classes

^a *Confidentiality* row does not sum to 100% because one scenario in this category did not meet the criterion to fall within high agreement or high disagreement classes.

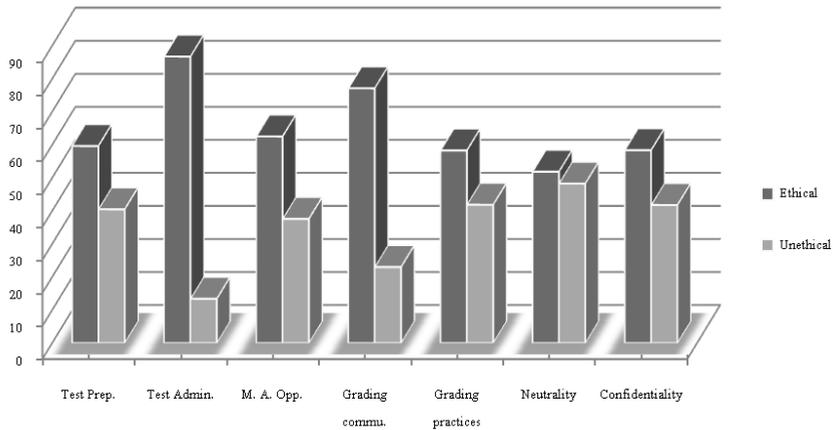


Figure 1: Percentage of teachers indicating the ethicality/unethicality of assessment practices in the seven categories

5.5 Why Did They Choose What They Chose?

In order to understand the existing dissimilarity among participants' judgments over the (un)ethicality of the scenarios, one item showing the least agreement among participants was selected within each category. Then, a minimum of 2 teachers for each category, who chose the item differently, i.e. one ethical and the other unethical, were invited for a brief interview to speak up their reasoning behind their choices.

In the first category, test administration, item 25 was believed to be ethical by 78.9% and unethical by 21.1% of the participating teachers. Those who thought it to be an ethical practice mentioned that writing an essay is not a single-shot one-session possibility. To them, it is not ethical to ask students to sit at their chairs in the final exam and write a comprehensive all-inclusive essay or they cannot

demonstrate their writing skill efficiently. On the other hand, those teachers who perceived this scenario to be unethical mainly stated their problem with the reliability of the test. They believed the essay brought to the exam session may be written by anyone other than the students and that they have to prove their writing skill under an equal situation with all the other students. It seems that the first group has more of a humanistic view toward assessment while the latter are concerned with the soundness and reliability of the tests. The problem occurs when one needs to make a choice between these two.

Under grading communication category, item 12 was checked as ethical by 46.3% and unethical by 53.7% of the teachers, holding the highest disagreement in the category. Those teachers who believed it would be ethical if they told students what materials were important to learn in preparing for a class test justified their choice stating that students need to learn what they are supposed to learn and there is no point in reading and learning unimportant or irrelevant parts of the course materials. However, teachers in the opposing group referred to the violation of educational aims. They clarified that education is not teaching 'to the tests' and that testing is only one part of a big framework of education. Quite like teachers in the previous comparison, participants in the latter group are mainly concerned with reliability of education and testing standards.

When it comes to confidentiality, less disagreement exists. 70.5% of participants believe that it is ethical to pass out scored tests to students in order of points earned, from the top score to the bottom score. Teachers on the ethical side stated that this strategy can increase their motivation in the following tests. Quite contrary to them, the remaining 29.5% of teachers who checked this scenario as unethical viewed this item 'very discouraging', especially on the part of the students who are usually scored low. In fact, as the

interviewee confirmed, “education should help students establish cooperation among themselves rather than competition.” Speaking of ‘Do Not Harm’ principle, it is interesting to notice that a majority of teachers have viewed this item as ethical. As it is clear, this behavior never protects individuals’ right (Green et al., 2007).

Item 27, in multiple assessment opportunities, attracted 46.05% of the participants to check it as an ethical practice, while 53.95% of them thought quite the other way around, which reveals a high degree of discrepancy among teachers. In this scenario, however, difference in opinion has to do with the way teachers perceive their career. Teachers in the unethical group stated that they could not find adequate free time and required resources to develop, administer, and score tests other than multiple choice items. The interviewee stated that “ethics of learning [for students] is not detached from that of teaching [for teachers]”, meaning that assessment ethics must be evaluated considering all the relevant issues including teachers’ concerns. To this scenario, on the other hand, the other interviewee replied differently. As he stated, no matter what, teaching implies adherence to a set of principles, such as ethics. In this regard, teachers are professionally and ethically obliged to provide multiple assessment opportunities for their students. This can cater for all the students with diverse proclivities (Eisner, 1994).

Grading practice, also, includes scenarios with grave disagreements among teachers. Out of 108 teachers, 41.4% stated that it would be ethical if a teacher increased a student’s final grade because he/she knew the student had mastered the course objectives in spite of his/her failure to complete some assignments (item 3). The interviewee, in this regard, justified his decision by criticizing the current assessment practices. He stated that, “We don’t have enough means to assess students’ degree of improvement, so it is quite evident that teachers use their personal understanding of it.”

When asked about his decision to choose the item related to the grading practice scenarios as unethical, the other interviewee preferred to 'stick to the rules'. To him, "we cannot appropriately and sufficiently examine each student's case and maybe we are not allowed to, lest it may lead to our subjective decreasing or increasing the scores." He believed that teachers should be as detached from students as possible: the more detached, the more objective.

When it comes to test preparation scenarios, 59.6% of the participants thought of it as an ethical assessment practice to provide students with a number of questions some ratio of which will be given in the exam. The interviewee in this group justified his choice by resorting to the lack of time and importance of testing excuses. As he put, "we are very limited in class with only 2 hours a week and, yet, we have to get students prepared for their final exam. To respect students' needs, I think, this teacher is doing good to prepare her students for the exam this way." The other interviewee, however, said he was concerned with the education no matter what the results may be. To him, "although it is good to help students with their final exams, teachers should not forget and ignore the major objectives of the course specified by the curriculum." In other word, in terms of test preparation, teachers seem to be compressed by the dilemma of observing the ethics of students and curriculum.

Finally, the last scenario presented to the interviewees was the one which has to do with students' identity, in neutrality category (item 18). To 41.5% of the participants, it is an ethical behavior to know which student wrote which essay. Quite like item 3 in grading practice category, the interviewee of this category stated that due to the lack of sufficient means to figure out students' improvement, a teacher should know how the performance of each student was. Also, the rest of participants (58.5%) were concerned about the possible subjective judgment on the part of teachers.

Ethical assessment practices in English classes

To conclude, interviewees' justifications can be divided into three classes. Interviewees responses were meant to observe the ethics of a. students, b. education, and c. teachers. Nevertheless, teachers do not view their importance equally. Table 13 summarizes the reasons participants prioritized for their choices.

Table 13: Interviewees' reasons for their choices

Students	Education	Teachers
Process-based Evaluation	Test Reliability	Lack of Required Time and Resources
Teaching to the Test (2 ^a)	Educational Objectives (2)	-
Providing Motivation	Teachers' Career Obligations	-
Providing Cooperation	Objectivity (2)	-
Multiple Assessment Oppor. (2)	-	-

^a Numbers indicate the frequencies above 1 for the stated reasonings

The following graph, also, provides a vivid understanding of the three classes of interviewees' reasons and their relevant percentages.

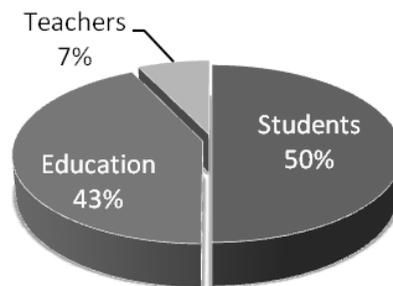


Figure 2: Percentages of interviewees' reasons

As shown in Figure 2, justifications the eight interviewees stated for their choices are generally classified into three classes: half of the

stated reasons (50%) are related to the students, 43% talked about issues dealing with education, and only one of these reasons (out of 14, i.e. 7%) was stated to refer to subjects related to teachers. The importance of this graph and its data is inherent in the fact that now we can claim, with 93% confidence, that teachers are significantly inclined toward and willing to cater ethics, whatsoever they view it, for the sake of students and education. This finding establishes a firm base for taking further steps: we know that teachers are inclined in the right way for prioritizing students and education, and though there is a great discrepancy among them in terms of the way they consider classrooms behaviors as ethical/unethical. Further research is needed to answer hows of internalizing ethics in teachers' assessment practices in particular and teaching practices in general.

6. Discussion

The present study was intended to reexamine the study carried out by Green et al. (2007) investigating the degree to which pre-service and in-service teachers viewed some assessment procedures as ethical or unethical. In line with the significance of the research mentioned in Green et al. (2007), the present study was an attempt to describe the status quo of English teachers' perception of ethicality or unethicality of some major assessment themes. Along with experience, gender, as an extension of Green et al.'s study, was also controlled.

Based on the in-depth analysis of the individual items and also examining the seven categories each taking in a number of scenarios spinning around and validated by a single theme, respondents' high agreement and high disagreement on different items were identified. Apart from detailed information elicited, the percentages of high agreement and high disagreement items are of special significance. As it is shown in table 12, respondents

revealed high agreement on 16 (out of 40) scenarios, i.e. 40% of all scenarios. This percentage has reached 57.5% (23 items) for high disagreement group which is not satisfactory (1 item, 2.5%, could not be located in the cited categories). A brief comparison of this data with that reported in Green, et al. (2007) shows that high agreement in the domestic context is 1.18 times lower and high disagreement is 2.3 times higher than that in the context reported. Three areas of investigation which stem from this comparison can be investigating reason(s) for such high disagreement, looking for long-run solution, and implementing short-run solutions.

The next point to highlight is construing the findings about the two variables of this study. As it was reported, statistically speaking, there was no significant difference found between males' and females' and novice and experienced teachers' ideas about the ethicality or unethicity of the assessment practices. Although it may sound quite positive, it is an alluding finding for one of the variables. As for gender, it *is* positive; it shows that gender-based bias is not prevailing, but when it comes to experience one expects much higher agreement among experienced teachers' attitudes compared with those of the novice ones, which is not the case. This, it is claimed, is a drawback in our assessment. The three steps mentioned above (tracing reasons, inventing long-run solutions, and applying short-run solutions) have to be implemented for this issue, as well.

The next point to ponder is the ranking order of the seven categories. As for the first three, *test preparation*, *communication about grading*, and *confidentiality*, high agreements are promising, but the next 4 categories, especially the last one, i.e. neutrality, with no agreed item, call for reconsideration of many affecting factors. For *multiple assessment opportunities*, for instance, Eisner (1994) discusses the necessity of implementing multiple opportunities for assessment simply because students hold idiosyncratic modes of

interpretation and creativity. Furthermore, the content analysis of the interviews with participants revealed that teachers are ready and willing to apply ethics in the proper way, yet they are in high incongruity among themselves about the way they view ethical practices. Consequently, teachers need to be educated about how to implement proper ethical assessment practices, how to analyze various grading practices, and how to prepare and grade students based on the right ethical deliberations.

7. Conclusion and Implications

The results of both quantitative and qualitative analyses indicated that ethical assessment, at least among English teachers in the country, is not at the level of consensus. Teachers come from a variety of cultural and educational backgrounds. Each teacher adds to this melting pot his/her own idea of ethicality and the results would be, one can say, justified ethical practices on the part of each teacher (Bullough, 2011). To put it differently, in the view of all teachers, their perceptions of (un)ethicality of the provided scenarios were righteously justified.

The reason seems to be twofold. For the first part, teachers are not educated to distinguish the right practice from the ethical one. During their teacher education courses, they were taught theories of language teaching, so they have become scholars in education; however, what was missing in their courses was provision of an insight as which practice *must* be applied irrespective of the seemingly unethical façade. For instance, the interviewee who added his “personal understanding” of students’ improvements had to be taught to behave objectively no matter what, so that possible subjective favoring would be eradicated. Consequently, there is a need to define and develop courses on ethical assessment/evaluation in teacher education courses (Mahony, 2009).

One the other hand, teachers who are employed by the ministry of education or private sectors are not provided with a set of principles nor a code of ethics in assessment, or even a simple list of to dos and not to does when confronted with ethical dilemmas in assessment. In fact, to the best of the researcher's knowledge, educational institutions do not provide their teachers with any sort of codes, be it for teaching or assessment, whatsoever. Now, teachers will have to act based on their instincts to solve such dilemmas, which itself leads to the current discrepancy in perceiving assessments ethical or unethical.

A very urgent need is felt for tracing the likely reasons for the existence of such great disparities among university teachers in terms of assessment practices and related ethical considerations. Research can be conducted on any areas of teacher education programs (TEPs), materials, and curriculum content (Eisner, 1993) to pinpoint the influencing factors. Next, long-run solutions must be sought, which can be linked to the previous query. Once the causes are identified, solutions can be followed. However, we also need to cure our current condition. As a result, short-run resolutions such as in-service educations may be fruitful. Scheire (2008) and Green, et al. (2008), among others, called for training educators and staff for proper assessment practices although teaching should be followed by actual implementation and revisions (Baumgart, 1996). This is a big gap in teacher training courses as well as university curricula and language departments for educating applied linguists in postgraduate programs who have been educated and are thus able to make ethical decisions in terms of classroom assessment.

Apart from the insights teacher education programs as well as curriculum designers can receive from this study, the findings can open a new area of research in the field, especially with reference to the professional community in Iran. Now that the lack of consistency in ethical assessment practices is duly noted, more

studies can shed further light on the issue. Researchers can include and control other relevant variables, such as field of the study, cultural background, EQ, and efficacy among others, in order to explore the status quo of ethical assessment practices among English (and non-English) teachers. Both longitudinal and correlational studies, also, can examine existence of possible relationship between ethical dilemmas in assessment and teachers' personal characteristics, like introversion and extroversion, and professional characteristics, such as pedagogical knowledge base. In the later stages, experimental studies will be due, too, having in mind the ethical considerations.

Acknowledgements

The author wishes to gratefully acknowledge Prof. Susan Green's and Dr. Nabizadeh's helpful comments and encouragements all through the study.

References

- Anderson, L. W. (2003). *Classroom assessment: Enhancing the quality of teacher decision making*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Antman, M. D. (2007). Assessment and grading in the beginning band classroom. Unpublished MA Thesis in Music Education, the Florida State University, US.
- Asmus, E. (1999). Music assessment concepts. *Music Educators Journal*, 86(2), 19-24.
- Avalos, B., & Aylwin, P. (2007). How young teachers experience their professional work in Chile. *Teaching and Teacher Education*, 23(4), 515-528
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Baumgart, N. (1996). Codes of ethics on assessment practice: Confronting a quandary or inculcating ethical character? Overview of two sample

Ethical assessment practices in English classes

- documents. *Assessment in Education: Principles, Policies & Practice*, 3(6), 393-396.
- Boyle, J. D. & Radocy, R. E. (1987). *Measurement and evaluation of musical experiences*. New York: Schirmer Books.
- Bullough, R. V., Jr. (2011). Ethical and moral matters in teaching and teacher education, *Teaching and Teacher Education*, 27(1), 21-28.
- Campbell, E. (1993). Strategic Leadership or Suspended Morality? How Principals Cope With Ethical Conflicts in Schools?. Paper presented at the XXI Annual Conference of the Canadian Society for the Study of Education, Carleton University, Ottawa, June 10, 1993.
- Cheng, L., Rogers, T., & Hu, H. (2007). ESL/EFL instructors' classroom assessment practices: purposes, methods, and procedures. *Language Testing*, 21(3), 360-389.
- Cole, N. C., & M. J. Zieky (2001). The new faces of fairness. *Journal of Educational Measurement*, 38(4), 369-382.
- Colnerud, G. (1997). Ethical conflicts in teaching. *Teaching and Teacher Education*, 13(6), 627-635.
- Eisner, E. (1994). *The educational imagination: On the design and evaluation of school programs* (3rd ed.). New York: Macmillan.
- Farrell, S. R. (1997). *Tools for powerful student evaluation* (2nd ed.). Milwaukee: Meredith Music Publications.
- Genesee, F., & Upshur, J. A. (1996). *Classroom-based evaluation in second language education*. Cambridge: Cambridge University Press.
- Gipps, C. V. (1995). *Beyond testing: Toward a theory of educational assessment*. London: The Falmer Press.
- Green, S., Johnson, R., Kim, D., & Pope, N. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education*, 23(7), 999-1011.
- Haladyna, T., Nolen, S., & Haas, N. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20(5), 2-7.
- Impara, J. C., Plake, B. S., & Fager, J. J. (1993). Educational administrators' and teachers' knowledge of classroom assessment. *Journal of School Leadership*, 3(5), 510-521.
- Jackson, P. W., Boosrom, R. E., & Hansen, D. T. (1993). *The moral life of schools*. San Francisco, CA: Jossey-Bass.
- Joint Advisory Committee. (1993). *The principles for fair student assessment practices for education in Canada*. Edmonton, Alberta. Retrieved February 16, 2010 from http://www.education.ualberta.ca/educ/psych/crame/files/eng_prin.pdf

- Mackey, A., Polio, C., McDonough, K. (2004). The relationship between experience, education and teachers' use of incidental focus-on-form techniques. *Language Teaching Research*, 8(3), 301-327.
- Mahony, P. (2009). Should 'ought' be taught? *Teaching and Teacher Education*, 25(7), 983-989.
- McNamara, T. (2001). Language assessment as social practice: challenges for research. *Language Testing*, 18(4), 333-349.
- Mehrens, W., & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless or fraudulent? *Educational Measurement: Issue and Practice*, 8(1), 14-22.
- Milroy, L., & Gordon, M. (2003). *Sociolinguistics: Method and interpretation*. Oxford: Blackwell.
- Melo, P. (2003). Ethical conflicts in teaching: The novice teacher's experience. *Connections*, (pp. 175-189). Retrieved February 16, 2009 from <http://www.educ.uvic.ca/Research/conferences/connections2003/12Melo102.pdf>
- Moore, W.P. (1993). *Preparation of students for testing: Teacher differentiation of appropriate and inappropriate practices*. Paper presented at the annual meeting of the National Council on Measurement in Education, Atlanta, GA.
- Oosterhof, A. (2001). *Classroom applications of educational measurement* (3rd ed.). Upper Saddle River, NJ: Prentice Hall, Inc.
- Pantic, N., & Wubbles, T. (2013). Teachers' moral values and their interpersonal relationships with students and cultural competence. *Teaching and Teacher Education*, 28(3), 451-460.
- Pope, N., Green, S. K., Johnson, R. L., & Mitchell, M. (2009). Examining teacher ethical dilemmas in classroom assessment. *Teaching and Teacher Education*, 25(5), 778-782.
- Popham, W. J. (2000). *Modern educational measurement: Practical guidelines for educational leaders*. Needham, MA: Allyn & Bacon.
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, 18(4), 429-462.
- Scheire, M. A. (2008). Mainstreaming process evaluation: Ethical issues in reporting interim results. In M. Morris (Ed.). *Evaluation ethics for best practice: Cases and commentaries* (pp. 144-154). New York: The Guilford Press.
- Scherff, L. (2008). Disavowed: The stories of two novice teachers. *Teaching and Teacher Education*, 24(5), 1317-1332.

Ethical assessment practices in English classes

- Schmeiser, C. B. (1995). Ethics in Assessment. *ERIC Digest*. Retrieved February 4, 2009 from <http://www.ericdigests.org/1996-3/in.htm>
- Shapira-Lishchinsky, O. (2011). Teachers' critical incidents: Ethical dilemmas in teaching practice. *Teaching and Teacher Education*, 27(3), 648-656.
- Shortt, D., Hallett, F., Spendlove, D., Hardy, G., & Barton, A. (2012). Teaching, morality, and responsibility: A structuralist analysis of a teachers' code of conduct. *Teaching and Teacher Education*, 28(1), 124-131.
- Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice*, 18(1), 23-27.
- Strike, K., & Soltis, J. (1985). *The ethics of teaching*. New York: Teachers College Press.
- Strike, K., & Soltis, J. (1992). *The ethics of teaching* (2nd ed.). New York: Teachers College Press.
- Taylor, K., & Nolen, S. (2005). *Classroom assessment: Supporting teaching and learning in real classrooms*. Upper Saddle River, NJ: Pearson Education, Inc.
- Watzke, J. (2007). Foreign language pedagogical knowledge: Toward a developmental theory of beginning teacher practices. *The Modern Language Journal*, 91(1), 63-82.
- Webster's Ninth New Collegiate Dictionary* (1988). Springfield, MA: Merriam-Webster.
- Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership*, 46(7), 41-47.

