

Testing Writing in the EFL Classroom: STUDENT EXPECTATIONS

NO EFL PROGRAM CAN DENY OR IGNORE THE SIGNIFICANCE OF TESTING FOR evaluating learners' acquisition of the target language. An important area of concern in testing is how students view their own achievements. Often students' expectations of test results differ from actual results. Students' grade expectations are often higher, which may negatively affect student motivation. This situation calls for raising students' awareness of their abilities.

The focus of this article is testing writing in the EFL classroom. Specifically, it describes a study comparing students' expectations of grades with their actual grades earned for essays assigned in Freshman English classes at the Lebanese American University. The results confirm a divergence between expected and actual grades, as has been reported in other research. The article concludes with implications for classroom teaching and testing.

Background

Experience has shown teachers, researchers, and school administrators that, just like language itself, testing practices in ELT are not static but dynamic and changing. One controversial area is testing writing, which requires that test construction and evaluation criteria be based on course objectives and teaching methodologies. In the English language classroom, especially at the high school and university levels, teachers are always challenged by how to reliably and validly evaluate students' writing skills, so that the students will be better prepared for internal and external proficiency and achievement exams. Indeed, writing in the academic community is paramount; a student can't be successful without a certain level of academic writing proficiency.

Another question that many ELT programs are addressing is how do students perceive the process used to evaluate their work? Do they know how they are being tested and what is acceptable by the standards of the institution and their teachers? These are questions this study seeks to answer, but first, it is necessary to differentiate between assessment and evaluation of writing and to present the main issues involved.

Assessing and evaluating writing

There are many reasons for testing writing in the English language classroom, including to meet diagnostic, proficiency, and promotional needs. Each purpose requires different test construction (Bachman 1990, 1991; Pierce 1991). Recent approaches to academic writing instruction have necessitated testing procedures that deal with both the process and the product of writing (Cohen 1994; Connor-Linton 1995; Upshur and Turner 1995). It is generally accepted by teachers and researchers that there are two main goals of testing: first, to provide feedback during the process of acquiring writing proficiency (also referred to as *responding* or *assessing*), and second, to assign a grade or score that will indicate the level of the written product (also referred to as *evaluating*).

The present study focuses on evaluating student essays, that is, assigning scores in order to indicate proficiency level. Evaluation of writing in ELT has a long history, with various procedures and scoring criteria being revised and adapted to meet the needs of administra-

tors, teachers, and learners (see Oller and Perkins 1980; Siegel 1990; Silva 1990; Douglas 1995; Shohamy 1995; Tchudi 1997; Bacha 2001). For testing writing, reliability and validity, as well as choice of topics and rater training, are important and must be addressed whatever the purpose of the testing situation may be (Jacobs et al. 1981; Kroll 1990; Hamp-Lyons 1991; Airasian 1994; Kunnan 1998; Elbow 1999; Bacha 2001).

Reliability

Reliability is the degree to which the scores assigned to students' work accurately and consistently indicate their levels of performance or proficiency. Correlation coefficients of .80 and above between readers' scores (inter-rater reliability) as well as between the scores assigned by the same reader (intra-rater reliability) to the same task are considered acceptable for decision making (Bachman 1990). There is research that indicates that the gender, background, and training of the reader can affect the reliability of scores (Brown 1991; Cushing-Weigle 1994). Thus, to maintain reliability many programs put heavy emphasis on the training of raters and as a result have obtained high positive correlations (Jacobs et al. 1981; Hamp-Lyons 1991).

Validity

Validity is the degree to which a test or assignment actually measures what it is intended to measure. There are five important aspects of validity (Hamp-Lyons 1991; Jacobs et al. 1981):

1. *Face validity* Does the test appear to measure what it purports to measure?
2. *Content validity* Does the test require writers to perform tasks similar to what they are normally required to do in the classroom? Does it sample these tasks representatively?
3. *Concurrent validity* Does the test require the same skill or sub-skills that other similar tests require?
4. *Construct validity* Do the test results provide significant information about a learner's ability to communicate effectively in English?
5. *Predictive validity* Does the test predict learners' performance at some future time? To what extent should we teachers com-

municate these reliability and validity concerns to our students? Teachers' awareness of the issues of reliability and validity is crucial, but perhaps equally important is how accurately students perceive their own abilities and the extent to which they understand what is considered acceptable EFL writing at the university level.

Perceptions of achievement

Research in how students perceive their language abilities compared with faculty perceptions and actual performance indicates that there is a problem that needs to be addressed (Kroll 1990). In a survey carried out by Pennington (1997) with students graduating from university in the United Kingdom, results indicated that 42 of the 48 students rated their writing ability as very good or quite good. In contrast, the teachers did not indicate such confidence. Another study indicated that first-year university students, who were L1 speakers of Arabic, rated their EFL writing skills in general as good, while faculty rated their skills as only fair (Bacha 1993). There were similar findings in another study comparing student and faculty grade expectations with actual test scores (Douglas 1995). In a needs analysis project carried out at Kuwait University, Basturkmen (1998:5) reported that "over 60% of faculty members perceived students to have inadequate writing skills." She also found that students' English language proficiency did not meet professors' expectations and students were not aware of the level of proficiency that was expected of them (Basturkmen 1998:5). Basturkmen concludes that one curricular objective should be to "raise students' awareness of the levels of proficiency which the faculty find acceptable" (1998:5).

If EFL students studying at the university level are deficient in academic language skills, a critical question is, to what extent are the students aware of their deficiencies? From the studies cited above, it appears they are not very aware of their deficiencies or, at best, seem to be more confident of their abilities—and thus hold higher grade expectations—than is warranted by their teachers' perceptions or by their actual test scores. This study will examine the problem in the Lebanese university context.

Survey on student grade expectations

Participants and procedure

During the Fall 2000 semester at the Lebanese American University, 150 students in the Freshman English 1 course in the EFL Program (the first of four required courses) were surveyed on their grade expectations. These courses stress essay writing and reading comprehension skills, focusing on sentences, paragraphs, and short essays. The students who completed the survey were L1 Arabic speakers who had studied English during their preuniversity schooling and were pursuing different majors in the Schools of Arts and Sciences, Business, Engineering and Architecture, and Pharmacy. They had English entrance scores equivalent to TOEFL scores of 525 to 574, and were enrolled in Freshman English 1 sections with between 25 and 30 students each.

Specifically, the survey was given in order to find out if there were any differences between students' grade expectations and the actual grades they earned. The survey was given two weeks before the end of the semester with the belief that students would have a better idea of their abilities later in the semester than they would at the beginning of the semester. They were requested to indicate the grade range they expected on two end-of-course essays. The five grade ranges were: below 60%, failing; 60–69%, fair; 70–79%, satisfactory; 80–89%, good; and 90–100%, excellent.

Essay 1 (E1) was given toward the end of the semester in the Freshman English 1 course. It is usually in the comparison or contrast rhetorical mode with a choice of different topics and completed in two fifty-minute class periods. During the first class period, students write a first draft. The teacher makes comments for improvement on the first draft, which is then rewritten during the second period. Essay 1 constitutes 20% of the final course grade.

Essay 2 (E2) was given at the end of the semester as part of the final exam for the course, which also included a reading comprehension and vocabulary component. The reading and vocabulary component of the final exam is similar in content for all Freshman English 1 sections, but students have a choice of three or four topics in the essay section with each topic requiring a different rhetorical mode. Essay 2 also constitutes 20% of the final course grade.

Table 2
Percentage of Students Selecting Each Grade Range for Essays 1 and 2
 Expected vs. Actual Grades (figures are in percentages)

	(90–100%)	(80–89%)	(70–79%)	(60–69%)	(below 60%)
Expected E 1	2.5	37.7	50.6	9.3	0.0
Actual E 1	0.5	4.0	41.6	42.1	11.9
Expected E 2	5.6	46.3	44.2	3.9	0.0
Actual E 2	0.0	6.9	36.1	42.6	14.4

The survey asked students to indicate their grade expectations for these two end-of-course essays. In addition, for each essay, the students were asked to indicate their grade expectations for the three major sub-skills of essay writing emphasized in the course: language (sentence structure, grammar, vocabulary, coherence, mechanics), organization (format, logical order of ideas, thesis and topic sentences), and content (major and minor supporting ideas). To indicate each expected grade, students selected one of the five possible grade ranges.

Results and discussion

A statistical comparison was made on a random sample of 30 surveys using the Wilcoxon Signed Ranks Test. This statistical test indicates whether there are any differences in mean ranks of scores when normal distribution is uncertain. Results of the Wilcoxon test indicated significant differences of $p < .001$ on all tests, confirming that the survey results showing differences between expected and actual grades are not according to chance and have a high degree of certainty.

It is not possible to pinpoint the accuracy with which individual students predicted their grades because the survey responses were tallied in mean averages. The results are most revealing when student expectations are examined as a whole and we can see that student grade expectations differed from actual grades.

Table 1 shows that the mean actual scores of the students on the two essays are one grade level lower (10%) than their mean grade expectations.

Since the gap between mean expected and mean actual grades is large, a whole proficiency level, a question raised is whether the students are aware of the criteria for each grade level. In other words, do students understand what is expected of them in the writing skills on which they are being tested? From random interviews with students and faculty, it seems they are not and that more work needs to be done in this area in the university's EFL program. All of our efforts to set up valid and reliable testing criteria seem self-defeating if the learners themselves are unaware of their potential achievement level or what is expected in their writing. These are important issues that need to be addressed in any educational program.

Table 2 compares the percentage of students who expected each of the possible grade ranges with the percentage of students who actually received those grades on Essays 1 and 2. We can see that no student expected to fail on either of the essays, but actual results show a failure rate of 11.9 percent on Essay 1 and 14.4 percent on Essay 2. The most accurate predictions were made in the grade range 70–79%. Perhaps many of the students placed their expectations in this range because it represented a cautious and modest expectation.

As can be seen in Table 2, expected and actual grades differed in the 60–69% grade range, with only 9.3% and 3.9% of the students accurately predicting grades on Essays 1 and 2, respectively. In the grade range 80–89%, students showed overconfident predictions of 37.7% and 46.3% on essays 1 and

Table 1
Differences in Mean Expected Grades and Mean Actual Grades
 (expressed as a percentage of total possible grade)

	Essay 1 (E 1)	Essay 2 (E 2)
Mean Expected Grade	74%	75%
Mean Actual Grade	64%	65%

Table 3

Percentage of Students Selecting Each Grade Range for Writing Sub-skills in Essay 1
Expected vs. Actual Grades (figures are in percentages)

	(90–100%)	(80–89%)	(70–79%)	(60–69%)	(below 60%)
Expected Language	7.1	36.1	44.1	12.7	0.0
Actual Language	0.5	3.4	36.9	36.5	22.7
Expected Organization	10.2	48.5	34.9	6.6	0.0
Actual Organization	0.0	4.4	40.9	41.9	12.8
Expected Content	9.0	49.1	36.4	5.6	0.0
Actual Content	0.0	5.9	38.9	45.3	9.9

2, while only 4.0% and 6.9% actually attained these levels, respectively. Students were most overconfident in their predictions of grades between 90–100%; only 0.5% of the students actually attained this score on Essay 1, and none did so on Essay 2.

Table 3 shows expected and actual grades for the three sub-skills of writing (language, organization, and content) in Essay 1 (E 1). It indicates that the actual scores were lower than student expectations and that failure was not expected. In fact, the findings show that for E 1 there is a failure rate of 22.7%, 12.8%, and 9.9% on language, organization, and content, respectively. Again, grade expectations and actual grades were closest in the grade range 70–79%. Students had much higher expectations than actually obtained for both of the upper grade ranges, 80–89% and 90–100%. Of the three sub-skills, language proved to be the weakest for students, indicating a need to focus more on this sub-skill in the classroom.

Table 4 shows expected and actual grades for the three sub-skills of writing in Essay 2 (E 2). Similar to E 1, it indicates that students' expectations in the sub-skills for that essay were higher than their actual test scores, and that all students expected to pass. In general, student expectations in the sub-skills were higher for E 2 than for E 1. Perhaps students gained more confidence in their abilities by the end of the semester and thus expected higher grades at the completion of the course, even though their actual scores do not support this expectation. In fact, no student attained a grade level of 90–100% in any of the sub-skills in E 2, and there were more actual scores in the failing range than in the grade range 80–89%. Also similar to E 1, students' expectations were most realistic in the grade range 70–79%.

Implications

The results obtained from this survey reveal that students and their instructors have differ-

Table 4

Percentage of Students Selecting Each Grade Range for Writing Sub-skills in Essay 2
Expected vs. Actual Grades (figures are in percentages)

	(90–100%)	(80–89%)	(70–79%)	(60–69%)	(below 60%)
Expected Language	9.5	38.0	45.7	6.8	0.0
Actual Language	0.0	5.9	34.7	42.1	17.3
Expected Organization	14.8	50.1	32.9	2.1	0.0
Actual Organization	0.0	6.9	36.1	44.6	12.4
Expected Content	10.1	50.4	35.3	4.2	0.0
Actual Content	0.0	7.9	37.1	42.1	12.9

ent perceptions of acceptable essay writing. This has important implications for writing evaluation in the university's EFL program. Teachers need to help students increase their awareness and understanding of the proficiency levels required in writing essays.

One way teachers can do this is by showing their students sample essays, perhaps drawn from the students' own work, that represent each of the grade levels from poor to excellent. These model essays could be photocopied for the class so that they can be read and discussed in detail. Students could take part in practice evaluation sessions by assigning grades for each sample essay, including the three sub-skills language, organization, and content, according to the criteria for essays used by the EFL program. Such practice evaluation could be done in small groups, with each group justifying the grades it assigns in short oral presentations to the rest of the class, followed by questions and discussion. Once this exercise is done, the teacher could discuss the different grade ranges and comment on the grades assigned by the groups in light of what grades the essays would likely receive in a testing situation.

A second way to raise students' awareness of essay evaluation criteria is through individual or small group conferences held periodically with the teacher. In fact, although student-teacher conferences are carried out irregularly, they have been quite successful in the EFL program at the university, especially for lower proficiency level writers. Students become more involved in the evaluation process and more aware of what is expected in their essays, and thus realistically build confidence in their writing.

In addition to these awareness-raising activities, teachers need to revisit periodically the writing criteria being used for essay evaluation in light of recent research and innovations in teaching writing. Teachers also might need to clarify criteria for the different proficiency levels for the various types of writing tasks assigned throughout a semester. Essay tests in certain rhetorical modes, such as narration or description, might require different evaluation criteria than those used for essays in the comparison or contrast mode. Although the essay tests included in this survey were from the end of the semester, teachers might want to consider whether they should evaluate

essays written earlier in the course according to objectives covered up to that point.

Conclusion

Testing is an inextricable part of the instructional process. If a test is to provide meaningful information on which teachers and administrators can base their decisions, then many variables and concerns must be considered. Testing writing is undeniably difficult. Although we teachers try hard to help students acquire acceptable writing proficiency levels, are we aware that perhaps our students do not know what is expected of them and do not have a realistic concept of their own writing abilities?

This article has reported the grade expectations of students and the actual grades they earned on two important end-of-semester essays. Results show that students' expectations are significantly higher than their actual proficiency levels. Developing test procedures for more valid and reliable evaluation is necessary and important; however, it does very little to motivate students to continue learning if their perceived levels of performance are not compatible with those of their teachers. In addition to the need to develop valid and reliable testing procedures, we must not overlook the need to raise students' awareness of their abilities. It is perhaps only through this understanding that genuine learning occurs.

Note: This is a revised version of a paper presented at the 21st Annual TESOL Greece convention, held in April 2000. The author received a grant from the Center for Research and Development at the Lebanese American University to support this research.

References

- Airasian, P. W. 1994. *Classroom assessment* (2nd ed.). New York: McGraw-Hill.
- Bacha, N. N. 1993. *Faculty and EFL student perceptions of the language abilities of the students in the English courses at the Lebanese American University, Byblos Branch*. Unpublished survey results, Byblos, Lebanon.
- . 2001. Writing evaluation: What can analytic versus holistic scoring tell us? *System*, 29, 3, pp. 371–383.
- Bachman, L. F. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- . 1991. What does language testing have to offer? *TESOL Quarterly*, 25, 4, pp. 671–672.

