

## Evaluating the Assessment Dimension of the Writing Component of a Preparatory English Program at a Public University: A Case Study\*

**Abdullah Coskun**

*School of Foreign Languages, Abant Izzet Baysal University*

*Email: coskun\_a@ibu.edu.tr*

**[Abstract]** The main objective of this study is to explore the assessment dimension of the writing component of a preparatory English program at a large English-medium state university in Turkey. English as a Foreign Language (EFL) instructors (N=50) and students' (N=414) opinions about the effectiveness of the assessment aspect of the program were obtained by means of questionnaires and interviews. The quantitative data were analyzed using Statistical Package for the Social Sciences (SPSS) 15 while the analysis of the qualitative data was done through content analysis. It was found that although instructors and students were generally pleased with how writing was assessed, some instructors argued that there is a need to standardize the assessment of the writing skills, organize workshops to train instructors for the standardization process, introduce double marking for the assessment of students' written works, use a more detailed assessment scale, include the writing tasks tested in the proficiency exam in the program, and incorporate writing portfolios. Likewise, some students believed that the grading was not very objective, and suggested that the assessment procedure applied to test students' writing ability in the exams should be standardized. These findings provided areas for improvement not only in the context of study but also in other similar EFL contexts.

**[Keywords]** English writing; assessment; evaluation

### Introduction

Writing is a productive skill enabling us to express thoughts in our mother tongue, to enhance our thinking skills and to become productive members of society (Grabe & Kaplan, 1997). Similarly, writing has an important place in learning a foreign language. Believing that writing is a vital skill both in first language and foreign language learning, Raimes (1983) emphasizes that writing aims to produce a connected text for a purpose and to a specific audience by means of the constant use of eyes, hands and the brain. For Raimes, there are three reasons why writing is an important skill in learning a foreign language. Firstly, while writing, students' grammatical structures and vocabulary are practiced and thus reinforced. Secondly, students learn to produce new ideas using the language and to take risks. Thirdly, students become engaged in the new language as they spend a lot of effort in the writing process.

As also suggested by Grami (2010, p. 8), "...based on the natural order hypothesis, writing is generally considered to be the language skill obtained last, but nevertheless it is as important as the rest." Regarding the importance of writing skills in English Language Teaching (ELT), the National Council of Teachers of English (NCTE) with 60,000 institutional and individual members published a report indicating that the writing skill is considered the key to achievement in both school and in the workplace (NCTE, 2004). It is also emphasized that language learners' academic achievement in content areas is generally dependent upon their ability to express themselves by means of writing (Mason, Benedek-Wood, & Valasa, 2009).

Despite the importance attached to EFL writing, it has been a historically neglected part of language programs (Badger & White, 2000). Due to the historical dominance of the audio-lingual approach in ELT (Matsuda, 2001, cited in Matsuda, 2003), studies carried out regarding EFL writing are profoundly insufficient. Furthermore, the concentration on writing instruction in EFL countries, such as Turkey has received little attention even though there have been a great number of studies about teaching English writing in English as a Second Language (ESL) contexts (Leki, 2001). As pointed out by Fujieda (2006), the number of studies dealing with writing in EFL contexts is limited; consequently, the field of L2 writing is still underdeveloped in the field of applied linguistics.

\* This study is based on part of my PhD dissertation entitled "Evaluation of the Writing Component of an English Language Teaching Program at a Public University: A Case Study"

Because the assessment of this important skill influences the effectiveness of writing programs in local contexts (Hamp-Lyons, 1991; Messick, 1996), it is suggested that assessing EFL writing skills, a multi-faceted process, should be studied in different contexts (Coombe, 2010). Offering some useful and practical guiding principles concerning writing assessment, the report prepared by the Committee on Assessment (2006) indicates that the best writing assessment practices should be informed by curricular goals, and that they are undertaken in response to local goals. Therefore, this evaluation case study aims to investigate the assessment dimension of a preparatory English writing program in a large English-medium state university from the perspectives of students and instructors. Such case studies carried out for evaluation purposes around the world have revealed areas for the improvement of English programs in different contexts (Beretta & Davies, 1985; Spada, 1987; Snow & Brinton, 1988; Sharp, 1990; Lynch, 1990; Murphy, 2000; Nam, 2005; Norris & Watanabe, 2007).

Likewise, there have been many program evaluation studies contributing to the quality of EFL programs of different universities in Turkey (Topçu, 2005; Güntek, 2005; Gerede, 2005; Örs, 2006; Karataş, 2007; Güllü, 2007; Tunç, 2010; Özkanal & Hakan, 2010). The evaluation studies carried out in Turkey dealt with various dimensions of specific EFL programs. However, to the author's best knowledge, no studies have particularly focused on the writing components of these programs or the assessment of the writing ability. Consequently, by evaluating the assessment component of the writing program in the context of this study, the aim is to bridge the research gap in the field, and to offer some guidance to the relevant stakeholders regarding the strengths/weaknesses of the assessment dimension of the writing program. Considering the purpose of this evaluation study, the following research questions have been formulated:

1. How do students evaluate the assessment dimension of the writing program at a preparatory English program?
2. How do instructors evaluate the assessment dimension of the writing program at a preparatory English program?

### **Assessment of the EFL Writing Ability**

As this study is concerned with the assessment dimension of an EFL writing program, there is a need for a concise discussion of a wide range of issues related to the skills assessed in tests, the elements leading to effective written assessment tests, the traditional and alternative assessment procedures, as well as reliability and validity in EFL writing assessment.

As it is the case for other skills, assessing writing is necessary to decide whether and to what extent students have achieved the goals of a language program. Emphasizing the complexity of assessing writing, Heaton (1989) suggests that the skills to be focused on when assessing a good piece of writing are mechanical skills, language use, stylistic skills, treatment of content and judgment skills. Before making a decision on the type of assessment to test these skills, some factors should be considered by teachers or test developers, including time available for assessment, cost of assessment, ease of assessment, and relevance of the dimension for the given task (e.g., content, rhetorical structure, organization, vocabulary, style, grammar, spelling, punctuation, accuracy of meaning) (Cohen, 1994). Similarly, Coombe (2010) draws attention to several questions to contemplate before the construction of a writing test. The first question is related to how much time should be allocated for a writing task. The second question is about whether all the stages in the writing process should be assessed, or only the product itself should be evaluated. The third question is pertaining to under which conditions the test should be administered. The fourth question is regarding whether all the students should write on the same topic or will they have the flexibility to choose from a variety of given topics. Another question is concerning who will rate students' written works, especially whether or not it should be the classroom teachers grading their own students' papers. Finally, the question as to whether more than one rater is a necessity to mark written pieces emerges.

In addition to these considerations, there are some elements to take into account so that effective writing assessment tests and tasks can be designed. As Hyland (2003) claims, one of these elements is the prompt, defined as the given instructions for a writing test or a task. Tribble (1996) suggests that these

prompts should be related to students' real-life problems, which keeps their level of motivation high. Additionally, it is important to provide students with prompts that will help them activate their declarative knowledge and prior knowledge.

Related to prompts, Coombe (2007) writes about two types of writing: guided writing and free writing. She differentiates between the two types by stating that tests of guided writing require students to manipulate content that is given through the prompt, generally in the form of a chart or diagram. On the other hand, free writing tasks expect students to understand a given prompt that describes a situation. In such tasks, writers formulate a response considering their prior knowledge, and their knowledge gained from the writing course. As far as free writing is concerned, Coombe also suggests that teachers should be clear about the expected form and length of response, specify the discourse pattern(s) the students are expected to use, and ask them to provide something beyond the prompt, such as an opinion. On the other hand, regarding the tests based on guided writing, Coombe proposes that multiple raters assess students' written works, agree on grading criteria in advance, familiarize students with the marking scheme by using it, teach sound writing strategies, and avoid controversial or cultural issues that might offend students.

Hamp-Lyons (2003) emphasizes that traditionally, the foreign language writing ability was assessed indirectly and objectively by means of multiple choice and grammar completion tests, which generally measure correct usage in sentence level constructions and the mechanics of writing (e.g., spelling and punctuation). Contrary to these traditional assessment tools, it is now believed that testing students' writing ability directly via more subjective test types is preferable to indirect testing instruments (di Ginnaro, 2006). Direct writing assessment requires students to produce the content by organizing the ideas and using appropriate vocabulary, grammatical structures, and syntax. Essay writing is an example of a direct writing assessment instrument, and essay tests are considered to be the most valid means of measuring foreign language writing ability (Schoonen, Vergeer, & Eiting, 1997).

However, as argued by Silva and Brice (2004), the timed direct essay tests cannot predict language learners' abilities to write under natural conditions, and as asserted by Qinghua (2010), there is dissatisfaction with timed, impromptu essay tests. Therefore, some alternative assessment tools have emerged. As reported by the members of the Minnesota Articulation Project (1997), students perform better when they are motivated by authentic reasons to use language. In their report, they emphasize the need for performance-based assessment that enables students to understand the development of their communicative competence. From their perspective, performance-based assessment is contextualized, authentic, task-based and learner-centered. As argued by Akar (2001), portfolio assessment is one of the most useful alternative performance-based assessment instruments commonly applied in the field of education for a variety of purposes. In general, portfolios include samples of evidence selected by the students (Brown & Hudson, 1998) so that teachers can realize the extent of learning and provide more individualized teaching. Students' projects, assignments, diary entries, self-assessments, comments from peers, and comments from teachers are examples of the written pieces that can be included in a portfolio (Hancock, 1994).

In a similar vein, Delett, Barnhardt, and Kevorkian (2001) underline the popularity of portfolio assessment to test students' writing ability. They believe that portfolios enable students to see their own progress over a certain period of time by means of the collection of their written pieces. Enginarlar (1994) illustrates the popularity of portfolio assessment among students in EFL writing classes in Turkey by revealing that they have favorable attitudes towards process writing and portfolio grading. He also emphasizes that portfolio assessment, which requires the assessment of students on the basis of several written pieces produced at different times, gives students the opportunity to revise their work and to receive feedback from their teachers as well as their peers. The implication of his study is that portfolio assessment in writing classes provides solutions to validity and reliability concerns, and it results in the beneficial backwash effect in the assessment of students' writing skills.

Self-assessment is another alternative means of testing writing. Some researchers favor the use of self-assessment procedures (Mesana, 2004; Liang, 2006). Self-assessment is believed to increase students' interest and motivation, positive attitudes towards writing, and consciousness of writing skills

and strategies (Mesana, 2004). According to Liang (2006), self-assessment involves students directly in their own learning processes and increases learner autonomy. Moreover, it trains students to assess their own language learning performance accurately, realize their weaknesses, and experiment with ways of self-correcting. Additionally, while raising students' knowledge about their learning goals and needs, self-assessment reduces the teacher's workload.

In tandem with assessment types, two different types of assessment scales have appeared: holistic and analytical. For Coombe (2010), holistic marking depends on the marker's general impression of the written work as a whole. When it comes to analytic marking, Weigle (2002) points out that raters might assess students' written pieces on different aspects, such as organization, register, vocabulary and grammar. Dwelling on the contextual factors in terms of the selection of assessment scales for a writing program, Coombe (2010) holds the idea that the teacher population, management structure of the institution, the availability of resources, and time can help the test developers make a decision about their marking procedures.

Finally, the issues of test validity and reliability have received a considerable amount of interest in the field of foreign language writing assessment. It is discussed in the relevant literature that while marking students' written pieces, inter-rater marking should be applied as a means to maintain test reliability (Kobayashi, 1992; Wood, 1993; Shi, 2001; Uysal, 2010; Coombe, 2010; Kayapınar, 2014). On the other hand, pertaining to test validity, it is thought that the tests should be consistent with what is taught in class (Morrow, 1991; Coombe & Evans, 2000; Wiggins & McTighe, 2005; Salend, 2009).

### **Context of the Study**

This study focuses on the assessment procedures applied in testing students' writing ability in a preparatory English program at a large English-medium state university in Turkey. As the name implies, preparatory English programs aim to prepare university students for the English medium education.

When students are eligible for enrollment at various English-medium universities based on their scores on the University Entrance Exam in Turkey, it is mandatory for them to be successful on an English proficiency exam administered by the preparatory English program of the university in which they are enrolled. Students' scores on this exam are taken into consideration to determine whether their English proficiency is at a level sufficient for them to follow their departmental courses in English. Those students who cannot get a passing grade in the exam have to attend a two-semester preparatory, intensive English program placing emphasis on general English skills, and then pass the English proficiency exam at the end of the second semester in order to be able to proceed to their English-medium departments.

In the context of this study, the main focus of the writing program is on paragraph development. The objectives of the writing program are as follows: writing simple sentences, writing compound sentences, writing complex sentences, using correct and appropriate vocabulary, using appropriate discourse patterns (cause-effect, compare-contrast, problem-solution, argumentation), writing academic paragraphs with appropriate topic, supporting and concluding sentences, achieving cohesion in writing via referencing, tapping into both new and prior information, achieving unity in writing through the use of appropriate signal words, and using appropriate register in writing (formal/informal language).

It was observed in the program that there is no single course book to teach writing, but in addition to handouts, the writing sections of the main course book used to teach general English in an integrated way are structured in writing-focused lessons. Handouts generally start with a brief introduction to the genre of writing to be covered in the lesson, and provide students with a model text so that they can see the organization of a typical example of the focused genre. Contained in the presentation of the materials there is a top-down process, in that after the model text is provided, linking words and useful expressions are supplied. This is followed by ideas for practice. Writing tasks are generally completed at home although the instructor may sometimes ask students to do the tasks in the class.

The achievement grade for each semester in the program is based on the scores students obtain from mid-terms, pop-quizzes and performance grades. The pop-quizzes generally assess students' reading, grammar and listening abilities. Students' writing ability is generally tested by three mid-terms in each term, and the writing sections make up twenty percent of the exams.

The English proficiency exam at the end of the year, on the other hand, contains sections such as language use, reading comprehension, listening comprehension and writing. The writing section makes up twenty percent of the total score and includes two tasks: responding to a reading text and writing a paragraph about a given topic.

### **Methodology**

Witte and Faigley (1983) assert that evaluation studies that are of a quantitative nature may only provide few insights concerning writing programs. They also warn researchers against the sole use of quantitative methods in evaluation studies involving not only students but also faculty, curriculum, and administrative structures. Therefore, a mixed-method research design involving both qualitative and quantitative data from students as well as teachers was used in this study.

### **Data Collection Instruments**

Both quantitative and qualitative data were collected from instructors and students for this study. Students and instructors completed a five-point Likert-type questionnaire (1=strongly disagree, 2=disagree, 3=not sure, 4=agree, 5=strongly agree), which also included an open-ended question. The student and instructor questionnaires related to the assessment dimension of the program were adapted from Erozan's (2005) evaluation study considering the context of the study. The questionnaires contained 12 closed items, through which participants were asked to express their views about the effectiveness of the way writing is assessed in the program.

While the closed-ended items in the questionnaires gather data about the perspectives of students and instructors regarding different aspects of the assessment procedures (e.g., 'Students were assessed on the things they practiced in the lessons'), an open-ended question at the end (e.g., 'Briefly list three changes you would recommend regarding how the writing performance should be assessed in the department') aims to obtain participants' suggestions for how the assessment dimension of the writing program could be improved. On the other hand, the face-to-face interviews with instructors were mainly based on the question 'Do you think that the assessment tools (e.g., mid-term exams) used for the assessment of the writing skills in the department are effective? If no, how should students' performance/success in writing be measured?'

### **Data Collection and Participants**

After the questionnaires for students and instructors in the program had been prepared, the items in the questionnaires were examined by the commission of the program coordinators and assistant chairs of the program by paying special attention to the comprehension difficulty and the suitability of the items in the context of the study. Additionally, five PhD-qualified experts in ELT, an expert in the field of Measurement and Evaluation as well as another expert in the field of Educational Sciences checked the questionnaires' content and face validity. The student questionnaire was piloted with 100 students, and the Cronbach's alpha coefficient of the questionnaire was found to be .81.

Approximately 15% of the entire student population at different proficiency levels in the program was given the student version of the questionnaire. Among 450 students, 414 students returned the questionnaire. Two hundred sixteen (216) of these students were male and 198 were female. The number of students between the ages of 17-19 was 322, and those between the ages of 20-22 consisted of nearly 22% of the participating students.

As far as the instructor version of the questionnaire is concerned, one third (N=50) of all the instructors teaching writing in the department (N=149) returned the instructor questionnaire. While 41 of these participants were female, the remaining 9 were male. In addition, 20 English instructors including the program coordinators volunteered to take part in the interviews or the open-ended surveys. The reason why open-ended surveys were used along with interviews was that an appropriate time for the interviews with some instructors could not be arranged due to their heavy teaching loads during the day.

### Data Analysis

The quantitative data collected via questionnaires were analyzed using SPSS 15 while the analysis of the qualitative data was accomplished by way of content analysis, which is based on the organization of data into similar categories, such as suggestions, concerns, strengths and weaknesses. According to Kiely and Rea-Dickens (2005), common patterns and trends emerging during the content analysis help researchers to better evaluate the language programs.

### Research Findings

The findings relevant to the first research question, “How do students evaluate the assessment dimension of the writing program at a preparatory English program?” are presented below. First, the results obtained from the closed and open-ended questions in the student questionnaire are discussed. Then, the findings from instructors' perspectives are illustrated on the basis of the second research question, “How do instructors evaluate the assessment dimension of the writing program at a preparatory English program?”

Table 1  
*Students' Opinions about the Assessment of the Writing Skills*

Items	Strongly Disagree		Disagree		Not Sure		Agree		Strongly Agree	
	N	%	N	%	N	%	N	%	N	%
1. We were assessed on the things (i.e. skills) we practiced in the lessons.	7	1.7	16	3.9	45	10.9	227	54.8	119	28.7
2. Before the tests (i.e. mid-terms), we were given information about the scope of the tests.	15	3.6	43	10.4	88	21.3	207	50	61	14.7
3. The directions in the mid-term exams were clear.	8	1.9	15	3.6	68	16.4	244	58.9	79	19.1
4. The materials used in the lessons and the test materials were similar in terms of difficulty level.	27	6.5	65	15.7	122	29.5	174	42	26	6.3
5. The mid-term exam results demonstrate my actual ability/proficiency in writing skills.	41	9.9	100	24.2	146	35.3	110	26.6	17	4.1
6. My writing skills have been correctly measured.	37	8.9	76	18.4	152	36.7	136	32.9	13	3.1
7. The grading was fair.	23	5.6	39	9.4	111	26.8	192	46.4	49	11.8
8. My performance in writing was good.	17	4.1	66	15.9	165	39.9	139	33.6	27	6.5
9. My writing skills have improved after the writing program.	11	2.7	26	6.3	105	25.4	216	52.2	56	13.5
10. I received sufficient feedback on my assignments.	13	3.1	29	7	91	22	208	50.2	73	17.6
11. I received sufficient feedback on my class work.	17	4.1	36	8.7	132	31.9	187	45.2	42	10.1
12. I received sufficient feedback on my performance in the mid-term exams.	13	3.1	41	9.9	130	31.4	186	44.9	44	10.6

The results presented in Table 1 revealed that the majority of the students evaluated some aspects of the assessment component of the writing program positively (‘agree’ and ‘strongly agree’) while a certain level of disagreement (‘disagree’ and ‘strongly disagree’) can be seen for certain items. For instance, as for positive responses with regard to the assessment of the writing skills, it can be stated that students expressed higher degrees of agreement (83.5% agreed or strongly agreed) with the item ‘We were assessed on the things we practiced in the lessons’. Many students (64.7% agreed or strongly agreed) felt positive about being informed beforehand about the scope of the tests. Even more students (78% agreed or strongly agreed) responded positively to the item regarding the clarity of the directions on the tests. In terms of students’ perceptions about their performance and progress in writing, 65.7% of the students

either agreed or strongly agreed that their writing skills have improved after the writing program.

Students expressed comparatively less agreement with some of the items, which can still be considered to be positively evaluated. For instance, approximately half of the students (48.3%) agreed or strongly agreed that the lesson materials and the test materials were similar in terms of difficulty level. As for the fairness of grading, 58.2% of the students expressed their agreement by marking either 'agree' or 'strongly agree'. When it came to the items pertaining to feedback on assignments, class work and mid-term exams; 67.8% of the students agreed or strongly agreed that they received sufficient feedback on their written assignments. On the other hand, nearly 55% of the students responded positively ('agree' or 'strongly agree') to the other related items, 'I received sufficient feedback on my class work' and 'I received sufficient feedback on my class work'.

As far as the comparatively negative points disagreed upon by a substantial number of students are concerned, it would be fair to say that a majority of students either disagreed or expressed uncertainty for certain items. While 22.2% of the students disagreed or strongly disagreed with the item, 'The materials used in the lessons and the test materials were similar in terms of difficulty level', 29.5% of the students indicated that they were not sure about it. The item, 'The mid-term exam results demonstrate my actual ability/proficiency in writing skills' was also negatively evaluated. About this very item, 34.1% of the students expressed disagreement and 35.3% students indicated uncertainty. It can also be stated that students held mixed feelings about the item, 'My writing skills have been correctly measured' as 27.3% of the students disagreed with it while 36.7% of the students indicated that they were not sure about it. Another item responded quite negatively by the students was related to the students' good performance in writing (40.1% agreed or strongly agreed and 39.9% marked 'not sure').

Additionally, students' responses to the open-ended question at the end of the questionnaire yielded some invaluable perspectives that could not have been found through the Likert-type items. Among 36 students who made comments about the assessment aspect of the program, there was a common concern shared by nine students about the 'subjective assessment' of their written pieces. In parallel with this concern, eight students proposed that two teachers should assess their paragraphs for more objective scoring. The same number of students focused on the mismatch between what was done in writing classes and what was tested in the writing part of the proficiency exam at the end of the academic year. These students suggested having more writing exercises which are relevant to the writing sections of the proficiency exam, and working on writing tasks similar to those tested in that exam (e.g., responding to a reading text). Complaining about the fact that there are only few writing pop-quizzes throughout the academic year, five students also commented that it would be a good idea to have more writing pop-quizzes. Furthermore, four students had objections to the limited time allocated to writing a paragraph in the exams. On the other hand, only two students suggested that their instructors should not penalize them for minor errors committed in the exams.

In brief, it would be accurate to state that students evaluated the assessment of the writing ability and their own performance in the program positively even though there were a few issues disagreed upon by most of the students. One of these issues was pertaining to the difference in the difficulty level of the lesson and the test materials. The other point was related to whether midterm exam results demonstrated students' actual writing ability.

Perhaps the reason why many students thought that the mid-term exam did not demonstrate their actual writing ability is because they held the belief that the grading was not very objective. It was suggested by these students that for reliable and objective scoring, two instructors should mark their written works instead of one instructor. The lack of relevance of the classroom activities to the proficiency exam was another concern shared by some of the students who recommended that writing tasks, such as responding to a reading text in a written form should also be focused on. Finally, some students underlined the need for more pop-quizzes and more time for writing tasks in the exams.

### ***Instructors' Opinions about the Assessment of the Writing Skills***

In addition to responding to the questionnaire items, 20 instructors including the program coordinators volunteered to make comments about the assessment dimension of the program through interviews. Some

of the volunteering instructors preferred writing their opinions on the open-ended survey because of their busy teaching schedule during the time of data collection. The quantitative findings obtained with regard to the assessment of the writing skills in the program are demonstrated in Table 2 below.

Table 2

*Instructors' Opinions about the Assessment of the Writing Skills*

Items	Strongly Disagree		Disagree		Not Sure		Agree		Strongly Agree	
	N	%	N	%	N	%	N	%	N	%
1. Students were assessed on the things (i.e. skills) they practiced in the lessons.	0	0	0	0	3	6	24	48	23	46
2. Before the tests (i.e. mid-terms), students were given information about the scope of the tests.	0	0	2	4	3	6	22	44	23	46
3. The directions in the mid-term exams were clear.	0	0	0	0	2	4	24	48	24	48
4. The materials used in the lessons and the test materials were similar in terms of difficulty level.	0	0	1	2	6	12	26	52	17	34
5. The mid-term exam results demonstrate students' actual ability/proficiency in writing skills.	3	6	6	12	23	46	13	26	5	10
6. Students' writing skills have been correctly measured.	1	2	4	8	22	44	16	32	7	14
7. The grading was fair.	2	4	3	6	7	14	28	56	10	20
8. Students' performance in writing was good.	1	2	6	12	17	34	21	42	5	10
9. Students' writing skills have improved after the writing program.	1	2	0	0	15	30	25	50	9	18
10. Students received sufficient feedback on their assignments.	0	0	0	0	4	8	34	68	12	24
11. Students received sufficient feedback on their class work.	0	0	0	0	8	16	31	62	11	22
12. Students received sufficient feedback on their performance in the mid-term exams.	0	0	1	2	13	26	24	48	12	24

The results presented in Table 2 illustrate that instructors generally expressed highly positive attitudes towards the assessment of the writing skills and students' performance in writing except for two items related to whether the mid-term exam results demonstrate students' actual ability in writing skills (marked 'not sure' by 46% of the instructors) and whether students' writing skills have been correctly measured (marked 'not sure' by 44% of the instructors). These items can be regarded as the least effective aspects regarding the assessment dimension of the program. It is worth mentioning here that students responded negatively to the same items.

On the other hand, among the issues evaluated positively by many of the instructors, those with the highest level of agreement ('agree' and 'strongly agree') were as follows: 'Students were assessed on the things (i.e. skills) they practiced in the lessons' (94%), 'Before the tests (i.e. mid-terms), students were given information about the scope of the tests' (90%), 'The directions in the mid-term exams were clear' (96%), and 'Students received sufficient feedback on their assignments' (92%).

Additionally, the open-ended question at the end of the questionnaire yielded some useful suggestions. Twenty-three (23) instructors answered the open-ended question about the improvement of the assessment dimension of the writing program. Out of 23, seven instructors thought that a standardization procedure should be established for the assessment of the students' written pieces. Similarly, five instructors argued for a more objective assessment system; in addition, according to four instructors, the writing assessment scale used in the program should be more detailed to ensure standardization among instructors. Regarding the assessment criteria, two instructors pointed out that students should be familiarized with the writing assessment criteria. Drawing attention to the mismatch between the proficiency exam and the writing activities focused on in writing lessons, six instructors suggested that there was a need to incorporate activities in line with writing tasks tested in the proficiency exam. According to three instructors, portfolio assessment should be a part of their assessment procedure

and from the perspective of two instructors, assessment ought to be based on regular written projects rather than timed paragraph writing in the exams. Moreover, two instructors recommended that, in the program there need to be workshops regarding issues, such as giving written feedback, written error correction, and the assessment of the writing ability. Finally, two instructors proposed that the word limit in the writing part of the exams should be increased, and students should be assessed on the quality of the writing rather than the quantity.

Moreover, the data collected through interviews and open-ended surveys conducted with 20 instructors yielded similar results. It was found that while more than half of these instructors (N=11) were generally satisfied with the assessment tools used for the writing component of the program, seven instructors shared their concerns that although the mid-term exams are fair enough to test students' writing ability, the proficiency exam is not. One of these instructors made the following comment with regard to the difficulty of the writing section and the type of writing required in the proficiency exam:

Midterms are fine because they test what we teach but the proficiency exam is far above their levels and they require free writing, which we never do in class. (Instructor 9)

Similar to the views of a few other instructors who made negative comments in response to the open-ended part of the questionnaire, seven instructors specifically touched upon the part of the writing section of the proficiency exam requiring students to respond to a reading text in the written form. One of these instructors made the following comment to argue that this part is not relevant to testing writing:

In the proficiency exam, there is a part, which requires students to comprehend a reading text and to respond to it. This part of the exam isn't fair as we are testing reading here, not writing. (Instructor 18)

Likewise, another instructor made comments about the part of the proficiency exam requiring students to write a paragraph about a given topic and underlined the dilemma between 'writing something well' and writing within 'certain types of discourse patterns':

[...] why should we try so hard to teach them certain types of discourse patterns? We should just let them write something well. Sometimes our students ask 'Why should I write it like this?' and they come up with some cliché sentences without really understanding these sentences. It looks very funny and mechanical. (Instructor 2)

On the other hand, four instructors urged for more writing quizzes and writing competitions among students while three instructors thought that the portfolio assessment could be an alternative to the timed paragraph writing in the exams. Moreover, six instructors suggested that the assessment criteria should be revised and there was a need for the standardization of the assessment of the writing skills. One of these instructors commented as follows:

Using the same grading scale, teachers can give very different scores to similar paragraphs. The assessment criteria should be more detailed. (Instructor 7)

Furthermore, one of these instructors made the following comment to compare the midterms and the proficiency exam in terms of reliability with a focus on the need for the standardization of the assessment of the writing skills, particularly in the midterms:

The problem is we don't have standardization of the assessment, especially at the midterms. The proficiency exam has a higher reliability rate than the midterms because the teachers who grade the writing don't know the students. For standardization purposes, we can make the teachers attend workshops. (Instructor 11)

Two instructors were also concerned about the unfairness of the assessment of the writing skills. Complaining about the scoring of the discourse patterns, one of these instructors said:

[...] most raters penalize students who do not follow these discourse patterns even if they answer the writing question correctly and even if they communicate their ideas. The same is true for the other way around. They give extra credits to those who use the expected discourse specific devices (linkers, enumerations, signaling words etc.) but do not really communicate ideas. (Instructor 16)

Overall, in terms of the evaluation of the assessment dimension of the program, it can be concluded that except for two items, instructors were generally satisfied with how writing was tested. The qualitative data, on the other hand, revealed that some instructors affirmed the need for more objective grading, and the establishment of a standardization procedure in marking students' written pieces. For a more standard scoring, some instructors argued for double marking of students' written pieces, a more detailed assessment scale, and the familiarization of the students with this scale. Encouraging instructors to attend workshops about the standardization of the assessment procedures was among other recommendations. Another important finding was that the proficiency exam included writing tasks that were not normally focused on in writing classes; therefore, some instructors proposed including certain writing tasks as tested on the exam (e.g., responding to a reading text in the written form) in the program. Finally, it was recommended that teaching discourse-level writing should be abandoned.

### **Discussion of Findings**

By means of quantitative and qualitative data collection instruments, this study aimed to investigate how EFL teachers and students evaluate the assessment component of a preparatory English writing program at a large English-medium state university in Turkey. Despite a high level of agreement with the effectiveness of the assessment procedures to test writing in the context of this study, some of the questionnaire items were not agreed upon by the majority of the participants. In addition to responses to closed-ended items in the questionnaire, some participants made comments and proposals. Although such proposals were voiced by a relatively low number of participants, the points they brought up hold considerable promise for the improvement of the assessment dimension of the writing program. To illustrate, some participants drew attention to the need for the standardization of the assessment. For standardization purposes, the need for double marking, checking inter-rater reliability, and a more detailed marking scale was underscored. The training of the raters by way of workshops was another suggestion. Furthermore, it was recommended that writing tasks that are tested on the proficiency exam and portfolios should be included in the program.

Nearly half of the instructors were not sure about the questionnaire items indicating whether mid-term exam results demonstrate students' actual ability in writing skills, and students' writing skills have been correctly measured. Likewise, many students disagreed with these items. These responses reveal the subjective opinions of the participants about certain aspects of the assessment procedures. However, as Mendelsohn (1989) points out, the perceptions of test-takers are very important in determining whether a test is good or not; therefore, after the test, test-takers should have the feeling that the test was fair and assessed what they needed to know. He also makes the following important remark to explain the relationship between students' perceptions of the test and their performance: "If test-takers do not find that a test meets their expectations, or they find that this kind of test is very strange to them, it can cause them to perform very badly" (p. 99). Consequently, these points negatively evaluated by both the students and teachers in this study should be improved so that students' writing performance can increase.

Moreover, some participants brought up the idea of double marking of students' written products. The literature supports this finding. For instance, Coombe (2010) underlines the importance of inter-rater reliability as a means to maintain writing test reliability. For the standardization of the assessment procedure for the writing ability, Kayapınar (2014) also believes that there is a need for double rating to

ensure a satisfactory level of inter-rater reliability. The necessity for multiple raters was also voiced for the assessment of the writing ability in the IELTS (International English Language Testing System) (Uysal, 2010).

Additionally, many researchers highlight the fact that raters may have biases against some students' performances, and their perceptions of good writing could influence their scoring (Kobayashi, 1992; Wood, 1993; Shi, 2001). For that reason, as recommended by a number of the participants in this study, raters should receive training in objective assessment. It is argued in the relevant literature that training raters can maintain consistency and fairness in EFL writing scores (Song & Caruso, 1996; Weigle, 2002; Johnson & Lim, 2009).

Similarly, for a more reliable assessment procedure, Enginarlar (1991) suggests the following three factors to take into account: established scoring systems, tested writing prompts and carefully trained raters. At the end of his study, he also presents some guidelines for scoring students' written pieces by means of a holistic marking scale. He recommends that instructors "should first read the papers quickly and place them in batches, ranging from 'very good' to 'nil' according to first impressions" (p. 43). The instructors should then read and reassess the written works again. In addition, he warns that "instructors should not grade papers for longer than one hour in one session (i.e., give a break when you are tired)" (p. 44). These guidelines can pave the way for a more reliable testing procedure in the context of the study.

On the other hand, some participants indicated that the marking scale used to assess students' written pieces in the program should be more detailed, and thus be improved. Herman, Aschbacher and Winters (1992) discuss some of the key features of an effective assessment scale. They argue that it should include a clearly defined list of criteria for the students to know what is expected from them. Underlining the importance of a marking scale in the effectiveness of assessment, McNamara (1996) also asserts that the score a rater gives to a piece of writing is the outcome of an interaction among the test, and a test-taker as well as the marking scale. Regarding the discussion about whether holistic or analytical marking scales should be used for the assessment procedure, Coombe (2010) points out that contextual factors, such as the management structure, the availability of resources and the amount of time allocated to the marking should be taken into account. More importantly, it is advised by Uysal (2010) that the raters' behavior while using the scales should be explored first by referring to a study carried out for IELTS by Micken (2003), who found that in spite of the use of analytic marking scales, raters had the tendency to respond to texts as a whole rather than to individual components. As a consequence, it would be fair to state that before making a suggestion about the type of marking scale to use in the context of the study, the raters' behaviors while using different scales ought to be studied.

It was also proposed by some participants that writing tasks tested on the proficiency exam should be covered in the program because the exam includes some parts (e.g., responding to a reading text in a written form) that are not focused on during the lessons. This finding leads us to the realization that the test should be consistent with what has been taught in class (Salend, 2009). Morrow (1991, p.111) also asserts that tests should "reflect and encourage good classroom practice". In the same vein, Wiggins and McTighe (2005) argue that instructional planning should link teaching and testing. It was also indicated by a few instructors in the program that there was no need to teach discourse-level writing to students. Uysal (2010) makes a similar recommendation for the improvement of the IELTS writing test by proposing that instead of strict rhetorical conventions, communicative aspects of EFL writing should be prioritized.

Furthermore, some participants felt the need for portfolio assessment in the program. According to the literature, the timed impromptu essay tests alone, as applied in the context of the study, cannot predict students' abilities to write under natural conditions (Silva & Brice, 2004; Qinghua, 2010). Coombe and Evans (2000) also state that instead of testing writing only on mid-term or final exams, students' writing ability should be assessed more frequently for a more fair and balanced assessment. Therefore, portfolio assessment ought to be implemented as an alternative assessment tool in the context of this study (Enginarlar, 1994; Delett et al., 2001; Akar, 2001; Apple & Shimo, 2004; Wang & Liao, 2008). It can also be maintained that portfolio assessment could be a viable solution to deal with validity and reliability concerns (Enginarlar, 1994) expressed by some participants in the study on condition that the necessary

training could be provided to writing instructors. Portfolio assessment is the product of the process approach considered by Matsuda (2003, p. 69) as “the most successful approach” based on the idea that writers follow steps, such as brainstorming about ideas to write, preparing drafts and revising the text in line with their teachers’ feedback.

### Conclusion

In conclusion, this case study revealed some points in need of improvement in terms of the assessment of the writing program. At the request of the program administrators, the findings of the study and relevant suggestions were shared with the Curriculum Core Committee, the research committee, the administration and instructors in the context of the study, and ways of making necessary changes in the program were discussed with them. During the discussion, it was agreed upon that the recommendations made in this paper are contingent on the organization's logistics, resources as well as available time.

As this study is only limited to the students and teachers in the English preparatory program in an English medium university, a comparative analysis of the same students’ perceptions when they are second or third year students can be made by future researchers to evaluate how well the program prepared students for their departments. Finally, as claimed by Erozan (2005), such evaluation studies revealing students’ and instructors’ suggestions for the improvement of a program can be considered as a way of identifying students and instructors’ needs. Therefore, a larger scale program evaluation study conducted with a larger student and instructor sample could add to the quality of the EFL writing program evaluated in this study.

Although this study uncovered results unique to the context of the research, the findings are believed to illuminate the common problems of other similar intensive EFL/ESL programs not only in Turkey, but also in other countries. Such contextual findings are considered helpful to the large body of practices in different parts of the world. As stated in the report prepared by the Committee on Assessment (2006), the methods and criteria to assess writing should be locally developed, deriving from the specific context and purposes for the writing being assessed. Also, as Cohen (1994) rightly argues, contextual factors are important in order to make decisions regarding the assessment dimension of the writing program. As a result, it would be fair to conclude that EFL programs should continuously get involved in such evaluation studies in their local contexts.

### References

- Akar, H. (2001). *Evaluating Higher Education Learners through Portfolio Assessment*. Informing Science 2001 Conference, June 19-22, Krakow, Poland.
- Apple, M., & Shimo, E. (2004). *Learners to teacher: portfolios, please! Perceptions of portfolio assessment in EFL classroom*. Proceedings of JALT Pan-SIG Conference (pp. 53-58). Japan.
- Badger, R., & White, B. (2000). A process genre approach to teaching writing. *ELT Journal*, 54(1), 153-160.
- Beretta, A., & Davies, A. (1985). Evaluation of the Bangalore Project. *ELT Journal*, 39(2), 121-127.
- Brown, J.D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653-675
- Cohen, A.D. (1994). *Assessing Language Ability in the Classroom*. Boston: Heinle & Heinle Publishers.
- Committee on Assessment Report. (2006). Retrieved from <http://www.ncte.org/cccc/resources/positions/writingassessment>
- Coombe, C. (2007). Fundamentals of language assessment. *Journal of NELTA*, 12(1&2), 36-46.
- Coombe, C. (2010). Assessing foreign/second language writing ability, *Education, Business and Society: Contemporary Middle Eastern Issues*, 3(3), 178-187
- Coombe, C., & Evans, J. (2000). Writing Assessment Scales: Making the Right Choice. *TESOL Arabia News*, 8(1), 7-9.
- Delett, J.S., Barnhardt, S., & Kevorkian, J.A. (2001). A framework for portfolio assessment in the foreign language classroom. *Foreign Language Annals*, 34(6), 559-568.
- di Ginnaro, K. (2006). Second language writing ability: Towards a complete construct definition.

- Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, 6(2), 1-17.
- Enginarlar, H. (1991). A quantitative and qualitative comparison of three techniques of grading ESL/EFL essays. *Journal of Human Sciences*, 10(1), 23-45.
- Enginarlar, H. (1994). Student attitudes to process-oriented portfolio grading. In C. Zaher (Ed.) *New Directions in Writing* (pp. 171-179) Cairo: Dar El Kutub.
- Erozan, F. (2005). *Evaluating the language improvement courses in the undergraduate ELT curriculum at Eastern Mediterranean University: A case study*. Unpublished doctoral dissertation, Middle East Technical University, Ankara, Turkey.
- Fujieda, Y. (2006). A brief history sketch of second language writing studies: A retrospective. *Maebashi Kyoai Gakuen College Journal*, 6, 59-72.
- Gerede, D. (2005). *A curriculum evaluation through needs analysis: Perceptions of intensive English program graduates at Anadolu University*. Unpublished master's thesis, Anadolu University, Eskişehir.
- Grabe, W., & Kaplan, R. (1996). *Theory and practice of writing. An applied linguistic perspective*. London: Longman.
- Grami, G.M.A. (2010). *The Effects of Integrating Peer Feedback into University-Level ESL Writing Curriculum: A Comparative Study in a Saudi Context*. Unpublished doctoral dissertation, Newcastle University, England.
- Güllü, S.A. (2007). *An evaluation of English program at Kozan Vocational School of Çukurova University: Students' point of view*. Unpublished master's thesis, Çukurova University, Adana.
- Güntek, D. (2005). *An Evaluation of the ENG 101, Development of Reading and Writing Skills 1*. Unpublished master's thesis, METU, Ankara.
- Hamp-Lyons, L. (1991). Reconstructing academic writing proficiency. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 127-154). Norwood NJ: Ablex.
- Hamp-Lyons, L. (2003). Writing teachers as assessors of writing. In B. Kroll (Ed.), *Exploring the Dynamics of Second Language Writing*. (pp. 162-189). New York: Cambridge University Press.
- Hancock, C.R. (1994). Alternative assessment and second language study: What and why? *ERIC Digest*. (Eric Document Reproduction Service No. ED 376695).
- Heaton, J.B. (1989). *Writing English Language Tests*. New York: Longman Inc.
- Herman, J.L., Aschbacher, P. R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Hyland, K. (2003). *Second Language Writing*. Cambridge: Cambridge University Press.
- Johnson, J.S., & Lim, G. S. (2009). The influence of rater language background on writing performance assessment. *Language Testing*, 26(4), 485-505.
- Karataş, H. (2007). *Yıldız Teknik Üniversitesi Modern Diller Bölümü İngilizce II dersi öğretim programının öğretmen ve öğrenci görüşlerine göre CIPP modeli ile değerlendirilmesi*. Unpublished master's thesis, Yıldız Teknik University, Istanbul.
- Kayapınar, U. (2014). Measuring essay assessment: Intra-rater and inter-rater reliability. *Eurasian Journal of Educational Research*, 57, 113-136.
- Kiely, R., & Rea-Dickens, P. (2005). *Program evaluation in language education*, New York: Palgrave Macmillan.
- Kobayashi, T. (1992). Native and nonnative reactions to ESL compositions. *TESOL Quarterly*, 26(1), 81-111.
- Leki, I. (2001). Material, educational, and ideological challenges of teaching EFL writing at the turn of the century. *International Journal of English Studies*, 1(2), 197-209.
- Liang, J. (2006). *Overview of self-assessment in the Second Language Writing Classroom*. 2006 TESOL Convention, Tampa, Florida. Retrieved from <http://www.secondlanguagewriting.com/documents/overview.doc>
- Lynch, B.K. (1990). A context-adaptive model for program evaluation. *TESOL Quarterly*, 24(1), 23-42.
- Mason, L.H., Benedek-Wood, E., & Valasa, L. (2009). Quick writing for students who struggle with

- writing. *Journal of Adolescent and Adult Literacy*, 53, 313-322.
- Matsuda, P.K. (2001). Reexamining audiolingualism: On the genesis of reading and writing in L2 studies. In D. Belcher & A. Hirvela (Ed.), *Linking literacies: Perspectives on L2 reading-writing connections* (pp. 84-105). Ann Arbor: University of Michigan Press.
- Matsuda, P.K. (2003). Second language writing in the twentieth century: A situated historical perspective. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 15-34). New York: Cambridge UP.
- McNamara, T.F. (1996). *Measuring second language performance*. London: Longman.
- Mendelsohn, D. (1989). *Testing should reflect teaching*. *TESL Canada Journal*, 7(1), 95-108.
- Mesana, C. (2004). An integrated approach to foreign language writing instruction. *Australian Review of Applied Linguistics*, 27(1), 44-57.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256.
- Mickan, P. (2003). What is your score? An investigation into language descriptors from rating written performance. *IELTS Research Reports*, 5(3). Canberra: IDP: IELTS Australia.
- Minnesota Articulation Project Report. (1997). Retrieved from <http://www.carla.umn.edu/assessment/mlpa/pdfs/miniguide.pdf>
- Morrow, K. (1991). Evaluating communicative tests. In Anivan, S. (Ed.), *Current Developments in Language Testing*. RELC, Singapore, pp. 111-118.
- Murphy, D.F. (2000). Key Concepts in ELT: Evaluation. *ELT Journal*, 54(2), 210-211.
- Nam, J. (2005). *Perceptions of Korean college students and teachers about communication-based English instruction: Evaluation of a college EFL curriculum in South Korea*. Unpublished doctoral dissertation, Department of Philosophy, Ohio State University.
- National Council of Teachers of English (NCTE) Report. (2004). Retrieved from <http://www.ncte.org/positions/statements/writingbeliefs>
- Norris, J.M., & Watanabe, Y. (2007). *Roles and responsibilities for evaluation in foreign language programs*. Retrieved from <http://www.nflrc.hawaii.edu/evaluation/files/Roles%20and%20Responsibilities%20booklet.pdf>
- Örs, M. (2006). *An analysis of the preparatory students' attitudes toward the appropriateness of the preparatory school program at the University of Gaziantep*. Unpublished master's thesis, University of Gaziantep, Gaziantep.
- Özkanal, Ü., & Hakan, A.G. (2010). Effectiveness of university English preparatory programs: Eskisehir Osmangazi University Foreign Languages Department English Preparatory Program. *Journal of Language Teaching and Research*, 1(3). 295-306.
- Qinghua, L. (2010). The Impact of Portfolio-based Writing Assessment on EFL Writing Development of Chinese Learners. *Chinese Journal of Applied Linguistics*, 33(2), 103-115.
- Raimes, A. (1983). *Techniques in Teaching Writing*. Oxford University Press
- Salend, S.J. (2009). *Classroom testing and assessment of all students: Beyond standardization*. New Delhi, India: SAGE.
- Schoonen, R., Vergeer, M., & Eiting, M. (1997). The assessment of writing ability: Expert readers versus lay readers. *Language Testing*, 14(2), 157-184.
- Sharp, A. (1990). Staff/student participation in course evaluation: A procedure for improving course design. *ELT Journal*, 44(2), 132-137.
- Shi, L. (2001). Native- and nonnative-speaking EFL teachers' evaluation of Chinese students' English writing. *Language Testing*, 18(3), 303-325.
- Silva, T., & Brice, C. (2004). Research in teaching writing. *Annual Review of Applied Linguistics*. 24, 70-106.
- Song, B., & Caruso, I. (1996). Do English and ESL faculty differ in evaluating the essays of native English-speaking and ESL students? *Journal of Second Language Writing*, 5(2), 163-182.
- Snow, M.A., & Brinton, D.M. (1988). Content-based language instruction: Investigating the effectiveness of the adjunct model. *TESOL Quarterly*, 22(4), 553-574.
- Spada, N. (1987). Relationships between instructional differences and learning outcomes: A process-

- product study of communicative language teaching. *Applied Linguistics*, 8(1), 137-161.
- Topçu, T. (2005). *An investigation of the effectiveness of the theme-based curriculum in the 2003-2004 academic year at the department of basic English at Middle East Technical University*. Unpublished master's thesis, METU, Ankara.
- Tribble, C. (1996). *Writing*. Oxford: OUP.
- Tunç, Y. (2010). *An evaluation of the English language teaching program at Atılım University based on stakeholders' perceptions: A case study*. Unpublished master's thesis, METU, Ankara.
- Uysal, H.H. (2010). A critical review of the IELTS writing test. *ELT Journal*, 64(3), 314-320.
- Wang, Y.H., & Liao, H. C. (2008). The application of learning portfolio assessment for students in the technological and vocational education system. *Asian EFL Journal*, 10(2), 132- 154.
- Weigle, S.C. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.
- Wiggins, G.P., & McTighe J. (2005). *Understanding by design*, Alexandria, VA: Association for Supervision and Curriculum Development.
- Witte, S., & Faigley, L. (1983). *Evaluating College Writing Programs*. Carbondale: Southern Illinois UP.
- Wood, R. (1993). *Assessment and testing: A survey of research*. Cambridge: Cambridge University.