INSTRUCTORS' PERCEPTIONS
OF
ENGLISH 101 WRITING EXAM GRADING CRITERIA
USED IN
THE FACULTY OF ACADEMIC ENGLISH PROGRAM
AT
BILKENT UNIVERSITY

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

YEŞİM TARKAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE GRADE OF MASTERS OF ARTS
IN
ENGLISH LANGUAGE TEACHING

SEPTEMBER 2005

Approval of the Graduate School of Social Sciences

_____

Prof.Dr. Sencer Ayata
Director

I certify that this thesis satisfies all the requirements as a thesis for
the degree of Master of Arts.

_____

Prof.Dr. Wolf König
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully
adequate, in scope and quality, as a thesis for the degree of Master of Arts.

_____

Assist. Prof. Dr. Gölge Seferoğlu
Supervisor

Examining Committee Members

Assist. Prof. Dr. Gölge Seferoğlu	(METU, FLE)	_____
Assist. Prof. Dr. Ayşegül Daloğlu	(METU, FLE)	_____
Dr. Elif Şen	(BU, BUSEL)	_____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are original to this work.

Name, Last name: Yeşim Tarkan

Signature:

ABSTRACT

INSTRUCTORS' PERCEPTIONS
OF
ENGLISH 101 WRITING EXAM GRADING CRITERIA
USED IN
THE FACULTY OF ACADEMIC ENGLISH PROGRAM
AT
BİLKENT UNIVERSITY

Tarkan, Yeşim

MA, Program in English Language Teaching

Supervisor: Assist. Prof. Dr. Gölge Seferoğlu

September   2005, 79 pages

This study aimed at finding out ENG 101 instructors' perceptions of the common ENG 101 writing exam grading criteria used in the Faculty of Academic English at Bilkent University. The main purpose of the study was to see what the instructors perceive as the positive and negative attributes of the criteria.

Fifty five instructors were involved in the study. The data were collected through quantitative and qualitative data collection instruments. In order to collect data, the instructors were asked to fill in a questionnaire which

consisted of both closed-ended response items using a likert-type scale and open-ended response items. Close-response items provided quantitative data and the qualitative data were derived from open-response items and the second instrument was interviews held with six volunteered instructors. The results showed that the instructors were mostly satisfied with the criteria in terms of its overall effectiveness, bands, and match between the course writing objectives and the criteria. However, the main problem found was about the equal weighting of the categories and the participants' lack of belief that the criteria were applied in a standard way across the Faculty of Academic English program. Based on the findings, suggestions were made taking into consideration the instructors' comments and evaluations.

Keywords: writing assessment, perceptions, instructors, writing assessment criteria

ÖZ

BİLKENT ÜNİVERSİTESİ FAKÜLTE AKADEMİK İNGİLİZCE GELİŞTİRME BİRİMİNDE KULLANILAN İNGİLİZCE 101 YAZMA SINAVI DEĞERLENDİRME ÖLÇÜTÜ HAKKINDA ÖĞRETİM ELEMANLARININ GÖRÜŞLERİ

Tarkan, Yeşim

Yüksek Lisans, İngiliz Dili Eğitimi Programı

Tez Yöneticisi: Y. Doç. Dr. Gölge Seferoğlu

Eylül 2005, 79 sayfa

Bu çalışma, Bilkent Üniversitesi Fakülte Akademik İngilizce Geliştirme Birimi'ndeki ENG 101 öğretim görevlilerinin, ENG 101 dersi için kullanılan yazma sınavı değerlendirme ölçütü hakkındaki görüşlerini ortaya çıkarmayı amaçlamıştır.

Bu çalışmaya 55 öğretim görevlisi katılmıştır. Veriler, nitel ve nicel veri toplama araçları ile elde edilmiştir. Veri toplamak için hem likert ölçeğinin kullanıldığı kapalı uçlu soruların, hem de açık uçlu soruların bulunduğu anket kullanılmıştır. Kapalı uçlu sorulardan nicel veriler sağlanmış, nitel veriler ise anketteki açık uçlu sorular ile gönüllü olarak çalışmaya katılan öğretim görevlileri ile yapılan mülakatlardan elde edilmiştir. Sonuçlar, öğretim elemanlarının genel olarak ölçütten memnun olduğunu göstermiştir. Ancak, bulunan temel problemler, ölçümde kullanılan kategorilerin eşit olarak

puanlandırılması ve ölçütün bütün öğretim elemanları tarafından aynı şekilde kullanıldığı yönündeki güvensizliktir. Çalışmanın sonunda elde edilen bulgulara dayanılarak öğretim elemanlarının yorum ve değerlendirmeleri doğrultusunda önerilerde bulunulmuştur.

Anahtar kelimeler: Yazılı değerlendirme, görüşler, öğretim elemanları, yazılı değerlendirme ölçütü

To my mother …

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

TABLES

CHAPTER I

INTRODUCTION

1.0 Presentation

This chapter introduces the background to the study, the FAE program, ENG 101 course, new FAE writing assessment criteria, the purpose of the study, the research questions and the significance of the study.

1.1 Background to the Study

Writing skill is one of the most important components of learning a language since even constructing a single sentence shows how well a student has mastered the target language. It is one of the ways to prove that students have made progress in learning the new language since it is a productive skill and requires some deeper processing. The importance of the ability to write effectively has increased more "as tenets of communicative language teaching - that is, teaching language as a system of communication rather than as an object of study – have taken hold in both second-and foreign- language settings" (Weigle, 2002, p. x). As a result, the more important writing has become, the greater demand there is for valid and reliable ways to test writing ability, not only for classroom use but also as a predictor of future professional or academic success. In other words, assessing writing has an important role in every class where students are asked to write. Evaluating students' writing is quite a challenging task for

English teachers. Assessment of writing ability is of crucial importance not only for teachers but also for students since many important decisions are made on how well they communicate in writing and such decisions affect students' education and even their lives (William, 1996; Brown, 1996; White, 1994).

On the other hand, assessing students' writing is not an easy task since "examiners are required to make judgments which are more complicated than the 'right – wrong' decisions…" (Alderson, Clapham & Wall,1995, p. 107). Testing students' writing ability in a reliable, valid and fair way is very crucial and the success lies in being able to assess something subjective as objectively as possible.

Testing and assessing writing is challenging due to difficulties inherent in it. There are certain basic considerations in assessing writing such as task variables, test –taker variables, rater variables, and rating scales (Bachman & Palmer,1996). Assessing writing requires subjective judgments on the part of raters; thus, teachers' perceptions of writing assessment and writing assessment rating scales are important.

As mentioned, one of the points to be taken into consideration is the rating scale. As Park confirms, "one of the first decisions to be made in determining a system for directly assessing writing quality is what type of scoring procedure will be used" (2004, p.1). Although there are some others, three types of scoring procedures have been mainly discussed in the literature: Analytic, holistic and primary - trait (Bachman & Palmer, 1996; Weigle, 2002; Alderson, Clapham & Wall, 1995). All of them have

advantages as well as disadvantages when they are applied. Considering the facts mentioned above, many researchers claim that no test or composition scoring procedure is perfect. As Perkins (1983) also states, the thing to be done is trying to find the best way for the context one has as no test or scoring procedure is suitable for all purposes. Another point that he makes and which is important to keep in mind is that "Even with guidelines and set criteria, the analytical and holistic scoring schemes can produce unreliable and invalid test information" (1983, p. 666).

As it has been highlighted before, raters have utmost importance while assessing students' papers. As raters use rating scales for assessing writing performance, when designing an effective rating scale, raters' perceptions of writing proficiency and well- worded and comprehensive descriptors that represent the construct of writing ability should be used. (Lumley, 2002)

Moreover, as Wharton (2003), in her study where she aimed to define appropriate criteria for the assessment of Master's level TESOL assignments claims group participation in the development of assessment practices is invaluable for everyone to feel ownership of the results. She also invited course participants – teachers with at least 3 years experience- to comment on the usefulness or otherwise of the assessment criteria.

Considering the literature, this study investigates instructors' perceptions of the ENG 101 writing exam grading criteria which is used to assess students' academic writing skills in the final ENG 101 exam in the faculty of Academic English Program at Bilkent University.

1.2 Background to FAE, ENG 101 Course and ENG 101 Writing Exam Grading Criteria

1.2.1 The Faculty Academic English Program

The Faculty Academic English Program (FAE) at Bilkent University provides English support courses to students in their faculties and schools. The courses offered by the FAE units range from content-based, academic skills courses in the freshman year to graduate writing courses for MA and PhD students. In providing academic skills support to a wide range of students in diverse faculties, instructors in the FAE program work cooperatively to design meaningful courses which emphasize high standards of academic writing achievement through challenging materials, active classroom learning, individual tutorial support and extensive feedback on student productions. In addition, in order to meet the needs of specific departments, instructors often work closely with faculty staff. The current organization of the post-preparatory programs in Bilkent University was established in January 2003 after the teaming up and merging of the First Year English Program with post-preparatory programs in Bilkent University School of English Language (BUSEL). There are currently five FAE units, each with approximately 15 teachers responsible to a Head, grouped according to the faculties or schools which they serve.

1.2.2 ENG 101 Course

ENG 101 course, which students have to take as an obligatory course in their first year, aims to introduce students to an academic approach to thinking, reading, speaking and writing in an integrated, meaningful manner so that they are able to apply the skills learnt in their departmental studies. In addition, the ENG 101 course aims to further develop the students' linguistic accuracy and range in English. To this end, there are many objectives to be covered in ENG 101. These objectives are grouped under the headings as academic thinking, reading, discussion /presentation, writing, and linguistic accuracy and document formatting.

In this study the main focus will be on the writing objectives which included academic writing, linguistic accuracy and document formatting (see Appendix A).

1.2.3 ENG 101 Writing Exam Grading Criteria

As it has been stated above, FAE consists of different units which are namely;

- Faculty of Engineering and Faculty of Science Unit   (FAE-FE / FS)

- Faculty of Economics, Administrative and Social Sciences Unit

  (FAE – FEASS)

- Faculty of Humanities and Letters, Faculty of Art, Design and Architecture Unit   (FAE – FHL / FADA)

- Faculty of Business Administration, Faculty of Law Unit

  (FAE - FBA/FL)

- Faculty of Music and Performing Arts, School of Tourism and Hotel Management, and the Vocational Schools of Computer Technology, Office Management, and Tourism and Hotel Services Unit   (FAE - VTS/FMPA)

For each unit ENG 101 course objectives are the same. This fact leads to the need for a set and standardized criteria to be used in each unit in order to be fair to students while assessing their progress – in this context academic writing skill is focused on. In the past, each of the five units had different criteria and this situation resulted in inconsistencies in assessing students' performance and this was not something desired for the course ENG 101. To avoid this, the director of FAE felt the need for standard writing criteria across the units. Then, from each unit the writing criteria used for ENG 101 were taken and after many interviews with the heads of the departments and instructors, a new set of criteria was designed. Having finalized the new criteria, the new criteria were launched at the beginning of 2004-2005 academic year.

1.3 Purpose of the Study

This study has been designed to investigate the instructors' perceptions about the ENG 101 writing exam grading criteria (see Appendix B) used to assess students' academic writing skills in the final ENG 101 exam in FAE program at Bilkent University.

1.4 Research Questions

This study will specifically address the following research questions:

1. How do Eng 101 instructors perceive the common Eng 101 writing exam grading criteria in terms of the following dimensions;

- Overall effectiveness

- Categories

- Descriptors

- Participants' feelings about its application

2. How would instructors mark the paper when a student's paper matched the B band in two categories but merits a C- band in the other two?

3. What do Eng 101 instructors perceive as positive attributes of the common Eng 101   Writing Exam Grading Criteria?

4. What do Eng 101 instructors perceive as negative attributes of common Eng 101 Writing Exam Grading Criteria?

5. What are the participants' suggestions for improving the Eng 101 Writing Exam Grading Criteria.?

1.5 Significance of the Study

As the objectivity of writing assessment has been a major concern in both ESL and EFL instruction and testing, this study is of particular value for

Bilkent University Faculty of Academic English. To begin with, instructors' evaluation of the new criteria through qualitative measures will provide feedback about the criteria. As a result, some points may arise for assessing and revising the criteria.

On the other hand, the findings may back up the usefulness and effectiveness of the criteria. The findings may indicate if further training would be needed to standardize marking and expectations. Another point is that if it is found that the criteria have been successful (efficient / accurate / user-friendly), it will assist the management in promoting program-wide implementation. Last but not least, the criteria itself, may help many other institutions while assessing students' writings since this criteria can be used as a benchmark and each institution may devise their own.

CHAPTER II

REVIEW OF LITERATURE

2.0 Presentation

This chapter reviews the literature on the importance of assessing writing, reasons of difficulty in assessing writing, as well as different methods and scoring procedures for assessing writing, and studies done on writing assessment criteria.

2.1 Importance of Assessing Writing

As Weigle (2002) highlights, the assessment of writing ability is of critical importance. Employers, academic instructors and writing teachers need to make decisions about potential employees and students based on how well they can communicate in writing. Park agrees with this issue stating that "the evaluation of writing ability of L2 students has become increasingly important in recent years, because the results of such evaluations are used for a variety of administrative, instructional, and research purposes" (2004, p.1). However, assessing writing is not something easy because of many inherent factors involved.

2.2 Why is Assessment of Students' Writing Difficult and Complex?

As assessment of students' writing is influenced by many variables including the human raters, context, etc. it is almost impossible to come up

with a scoring tool which would be universally reliable and valid in every context for every piece of writing that one produces. There are many variables affecting the scores that students get when their writings are assessed. These firstly consist of the construct and can include the tool or scoring procedure used as well as the raters. In other words, to overcome such variables it is essential to have both valid and reliable rating scales and procedures and this has become a major issue recently. There may be different points of view among the professionals in the field about how writing should be assessed but the fact that validity and reliability need to be obtained to the maximum extent is not questioned. White (1994) regards reliability as a prerequisite for validity as he believes that one may have a perfectly valid final exam; that is; it may perfectly reflect what has been taught. However, if it is scored by throwing the papers down stairs, grades would be meaningless. What he means is that no assessment device can be more valid than it is reliable. To sum up, since both utmost validity and reliability is desired, it is quite challenging to assess students' writings objectively.

2.3 Different Methods Used for Rating Students' Writing Performance

The methods used for rating students writing performance are mainly divided into two; direct and indirect. In indirect methods students are not required to perform a composing process. In other words, in such tests students' knowledge of the rules and conventions of English is assessed. The examples of such tests given by Heaton (1988) are dictation, multiple

choice for spelling, completion items, multiple choice for style and register etc. These are easier to mark objectively, quickly and reliably.

Opposed to indirect methods, direct methods of assessing writing aim to test students' writing performance through production of writing. Recently, since such tests incorporate opportunities for students to analyze, generalize, synthesize and evaluate information, they have been favored. Many institutions today employ such tests because they measure certain writing abilities more effectively than indirect tests of writing. Moreover, they motivate students to improve their writing. Last but not least, it is easier and quicker to prepare than indirect tests of writing. What is more, the authenticity of such tests is worthwhile having because the majority of them are real life tasks.

Although direct tests seem to be far better, they have several drawbacks. In Heaton's (1988) point of view, the main problem lies in unreliability. This is due not only to the subjective nature of the scoring of direct tests but also students' different performance on different topics and on different occasions. On top of this, the raters tend to spend considerable time on marking.

2.4 Different Types of Scoring Procedures for Writing Assessment

Weigle (2002) claims that the scoring procedures are critical because the score is ultimately what will be used in marking decisions and in making inferences about writers. Defining the rating scale which would be chosen according to what is focused on or aimed at that specific task and ensuring

that raters use the scale appropriately and consistently must be the central consideration in scoring. This will ensure the reliability and validity of scoring. McNamara notes that "the scale that is used in assessing performance tasks such as writing tests represents, implicitly or explicitly, the theoretical basis upon which the test is founded; that is, it embodies the test (or scale) developers' notion of what skills or abilities are being measured by the test" (cited in Weigle, 2002, p.108). Therefore, the development of a scale and descriptors for each scale level are of critical importance for the validity of the assessment.

2.4.1 Types of Rating Scales

The very first thing to be taken into consideration in determining a system for scoring is what type of rating scale is to be used; that is, should a single score be given to each essay, or will each essay be scored on several different features?

There are mainly three types of rating scales in the literature; primary trait scales, holistic scales and analytic scales.

In this section, each of them will be defined briefly and the advantages and disadvantages of each will be discussed.

2.4.1.1 Primary Trait Scoring

As Weigle (2002) states primary trait scoring is most closely associated with the work of Lloyd-Jones (1977) for the National Assessment of Educational Progress (NAEP), a large scale testing program for schools in

the US. The philosophy behind this type of scoring is that it is important to understand how well students can write within a narrowly defined range of discourse (e.g. persuasion or explanation). Such a rating scale is defined with respect to the specific writing assignments and essays. Parallel to what Weigle says, Perkins states that "in primary trait scoring, a holistic score is assigned to a particular feature of writing such as structure, tone or vocabulary. This procedure requires that graders ascertain whether a piece of writing exhibits certain characteristics – termed primary traits – which are crucial to a specific rhetorical task a writer is trying to perform" (1983, p. 658).

In other words, we can conclude that primary trait marking seeks to be more focused than general impression marking by being tailored to the specific writing task. When we focus on the advantages of such scoring, it can be said that it allows students to focus on a specific task and apply structures to a rhetorical context. It also provides a precise, detailed description of a student's writing ability for a specific task.

However, Weigle (2002) highlights the shortcoming of primary trait scoring as being very time and labor intensive since a scoring guide must be developed for every writing task.

2.4.1.2 Analytic Scoring

As Weigle (2002) confirms, in analytic scoring, various features of a composition are focused on separately. This depends on the purpose of the assessment. Essays might be rated on such features as content, organization, cohesion, register, vocabulary, grammar and/ or mechanics.

The weighting of each section may differ according to the focus. Heaton (1988) suggests that the rater may need to alter either the components or the weightings given to them according to the proficiency level involved.

The primary advantage of analytic scoring is that it provides more useful, detailed information which helps to diagnose the students' writing abilities. Hughes (1989) adds that scorers are compelled to consider aspects of performance which they might otherwise ignore. Some other research suggests that analytic scoring is much more beneficial in rater training, as inexperienced raters can understand how to apply the criteria in separate scales more easily (Francis, 1977; Adams, 1981, cited in Wier, 1993). Last but not least, as the rater has to give a number of scores, the scoring can be more reliable.

On the other hand, the major disadvantage of analytic scoring is that it is much more time consuming as the rater has to make more than one decision for each piece of writing. The second disadvantage of such scoring is that concentration on different aspects may affect the attention and coherence may be lost. The last point to raise is that different raters may value different aspects of an essay.

2.4.1.3 Holistic Scoring

Holistic scoring aims to rate the overall proficiency level reflected in a given sample of students' writings. When holistic scoring is employed, each writing sample is read quickly and then judged against a rating scale, or scoring rubric that outlines the scoring criteria. The difference between

holistic scoring and general impression marking is the criteria which are never explicitly stated when general impression making is carried.

Holistic scoring has become widely used in writing assessment over the past 25 years because it has a number of positive features. The primary advantage is that it is faster to read an essay once and assign a single score than read it several times, each time focusing on a different aspect of the writing (Weigle, 2002). White (1984) maintains that holistic scoring is intended to focus the readers' attention on the strengths of writing so that writers are rewarded for what they do well. He also argues that holistic scoring is more valid as it reflects most closely the authentic and personal reaction of a rater to a text.

Similar to other scoring techniques, holistic scoring also has some drawbacks. Firstly a single score does not provide useful diagnostic information about students' ability to write. Second disadvantage that Weigle states "holistic scores are not easy to interpret, as the raters do not necessarily use the same criteria to arrive at the same scores: for example, a rater could give a 4 to a certain essay because of its rhetorical features and a second rater could give a 4 to the same paper because of its grammar and vocabulary" (2002, p.115).

The three scoring procedures, primary trait, analytic and holistic, are the ones mostly used while assessing writing. However, the choice about the kind of rating scale to be used is not always clear-cut. Therefore, institutions sometimes design their own criteria.

As Crusan (2002) states writing assessment has always been problematic since assessing writing is subjective and there are no perfect assessment criteria. Douglas adds that "There is very little discussion in the standard language testing literature about the provenance of assessment criteria" (2001, p. 173).Therefore, most educational program developers try to design their own scales based on certain features derived mainly from these three scoring procedures.

As Park claims, it should be noted that;

No test or scoring procedure is suitable for all purposes and even with the guidelines and set criteria, such procedures can yield unreliable and invalid test information. Therefore, decisions regarding the selection of evaluation procedures need to be made within the context of a specific testing situation (2004, p.3).

Turner also emphasizes the importance for educators to be aware of scale construction procedures because "this awareness can contribute to their understanding and interpretation of rating scale criteria"(2000, p.556).

Wharton (2003) also supports Turner by stating that participants should be actively involved in the process of defining criteria, so that commitment to the scale and consistent application of it can be achieved.

Hawkey and Barker (2004) in their study aiming at developing a common scale for the assessment of writing made use of expert and participants' reactions to draft scales in order to revise the rating scale.

Participants' involvement in designing, revising and applying rating scales is essential. However, recent studies have focused on the rating

process and participants' approach of assessing written work by asking participants to describe the factors that contribute to their final rating other than the scale at hand. Lumley (2002) found out that although raters understand the rating category in general terms, they sometimes apply the rating scale in various ways. Raters give different emphasis to the different components of the scale descriptors.

In another study, Cumming (1990) reports that of the 28 categories of comment that he identified, 20 could be classified under the three categories for which raters had to award scores. These were substantive content, language use and rhetorical organization.

When raters assess students' written products, raters' awareness of underlying construct that is being measured via assessment criteria and raters' perceptions of assessment criteria they are using are important. Each rater has their own, different perspectives and so the relationship between raters and assessment criteria calls for investigation.

CHAPTER III

METHOD

3.0 Presentation

This chapter of the study covers the participants, instruments and the procedures.

The aim of this study is to investigate the perceptions of the FAE instructors of the new common ENG 101 Writing Exam Grading Criteria in terms of several dimensions. To determine what ENG 101 instructors perceive as both the positive and negative attributes of the ENG 101 writing exam grading criteria has been the primary focus of the study. Lastly, as a result of all the findings, what suggestions towards improving the criteria could be made will be looked at.

3.1 Overall Design of the Study

The purpose of the study was to get Eng 101 instructors' perceptions about the ENG 101 writing exam grading criteria used in the FAE program at Bilkent University. The data were collected via quantitative and qualitative data collection instruments. The questionnaire designed provided both qualitative and quantitative data. The interviews held also provided further qualitative data. Fifty instructors were given the questionnaires and 6 of them were interviewed to get their perceptions on the criteria in detail.

3.2 Research Questions

The research questions in this study are;

1. How do Eng 101 instructors perceive the common Eng 101 writing exam grading criteria in terms of the following dimensions;

- Overall effectiveness

- Categories

- Descriptors

- Participants' feelings about its application

2. How would instructors mark the paper when a student's paper matched the B band in two categories but merits a C- band in the other two?

3. What do Eng 101 instructors perceive as positive attributes of the common Eng 101 Writing Exam Grading Criteria?

4. What do Eng 101 instructors perceive as negative attributes of common Eng 101 Writing Exam Grading Criteria?

5. What are the participants' suggestions for improving the Eng 101 Writing Exam Grading Criteria?

3.3 Participants

Fifty-five instructors out of 64 were involved in the study. Not all the instructors were involved as, during the administration of the questionnaire session, they were teaching at summer school and they could not attend the session. One of participants was the head of the FAE program. The other five are the heads of each unit and the rest are the instructors who give ENG courses to the students at the departments. Out of 55 instructors, 24 of them were male and 31 of them were female. Twenty- seven instructors were native and 28 were non-native. Four of the instructors had a PHD degree whereas 41 of them had a BA degree. Their experience in the FAE program ranged from 4 months to 16 years. Although during the administration of the questionnaire all 55 instructors seemed to be answering the questionnaire, when it was time for data analysis, it was noticed that 5 of them had just filled in the first section from which demographic data was gathered. This means they did not fill in the rest of the questionnaire stating that they had not used the ENG 101 writing exam grading criteria as they had not taught ENG 101 course since the new criteria was launched. As a result, the data analysis was done based on 50 instructors' responses.

3.4 Instruments

3.4.1 Questionnaire

In order to collect data on FAE instructors' perception of the ENG 101 Writing Exam Grading Criteria, a questionnaire was designed.

Questionnaires are widely used and useful instruments for collecting survey information, providing structures, often numerical data, being able to be administered without the presence of the researcher, and often being comparatively straightforward to analyze (Wilson & Mclean, 1994, cited in Cohen et al, 2000).

The questionnaire had 4 parts (see Appendix C). The first part asked for biodata, which is information about respondents' background and individual characteristics. The second part was made up of closed-response items using the likert scale. In this part, the likert scale was used as "likert scales are generally useful for getting at respondents' views, judgments, or opinions…" (Brown & Rodgers, 2002, p. 120). In this research a 1 to 4 scale was used (1- Strongly Agree, 2- Agree, 3- Disagree, 4- Strongly Disagree) as the respondents were expected to state their perceptions as positive or negative rather than being noncommittal.

Although closed-response items are mostly preferred in questionnaires as "they are quick to complete and straightforward to code and do not discriminate on the basis of how articulate the respondents are (Wilson & McLean, 1994, cited in Cohen et al, 2000), a box was added next to each item to enable the participants to write or make extra comments about each statement to express themselves further.

In the 3rd and 4th sections, there were open-response items where the participants could express their thoughts and opinions more freely in a detailed way.

Before administering the questionnaire, the items were written keeping some key points in mind such as things to avoid in writing good survey items (Brown & Rodgers, 2002; Bailey, 1994; Cohen et al 2000). Even though some questions can be seen as overlapping or repetitive, the aim by having such items or sections was to have 'reliability – check question pairs' (Bailey, 1994, p.134). In the second section, the likert scale was preferred as "rating scales are particularly useful for tapping attitudes, perceptions and opinions of respondents." (Cohen et al 2000, p. 255) The questionnaire consisted of not only a scale but also open ended questions as "a questionnaire might be tailored even more to respondents by including open-ended questions to which respondents can reply in their own terms and own opinions", (Cohen et al 2000, p. 255). All the items in the questionnaire and in the interview questions were grouped to get feedback from the instructors under certain categories – headings- which were namely;

- Overall effectiveness

- Categories

- Bands

- Descriptors

- Match between Eng 101 course writing objectives across the FAE program and the criteria.

- Suggestions for improvement

In the questionnaire, the first eight questions were designed to find out the overall effectiveness of the criteria. Items nine and ten were to get

instructors' opinions about the categories in terms of their weighting and match with the course writing objectives. The next four items aimed to have feedback specifically on descriptors in each category in each band. Finally, the last four items were asked to see how the instructors feel about the application of the criteria both by themselves and others across the program.

In the next section, section C, a scenario was given to find out how they use the criteria while marking. The aim here was to see if they apply the criteria in the same way or not while marking in the given situation.

Section D aimed to get instructors' perceptions on the criteria in terms of positive and negative attributes of them by asking them to identify the strengths of the criteria and the points to reconsider in the criteria.

The last section, section E, was designed to see what the instructors' would suggest to improve the criteria.

Then, as a next step the questionnaire was piloted. This was mainly to increase the reliability, validity and practicality of it (Oppenhaim, 1992; Patton, 1990; Brown & Rodgers, 2002). "A common way to do this is to have someone look at the content and format of the instrument and judge whether or not it is appropriate" (Fraenkel and Wallen, 2000, p.171). In this research the questionnaire was given randomly to some instructors to have a look and make comments regarding the clarity of the questionnaire items, instructions and layout without actually answering it. As well as the feedback from instructors, two experts from the field of English Language and Education

were also consulted during this stage. The aim for this was to check face and content validity of the instruments.

Having followed the key points while preparing a questionnaire (i.e. avoiding leading, complex, irritating questions negatives etc. (Oppenhaim, 1992; Brown and Rodgers, 2002; Patton, 1990; Bailey, 1994) and having made the necessary changes, the questionnaire was administered to 55 instructors.

After administering the questionnaire and entering the data into the SPSS program, the reliability of the questionnaire was found to be at the cronbach alpha level 0,91 which proves that its reliability is high (see Table 1)

Table 1 Reliability   Analysis of the Questionnaire

Item-total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
B1	38,4400	60,0065	,6891	,9112
B2	38,2600	58,0739	,6378	,9126
B3	38,4400	59,2718	,6391	,9122
B4	38,2600	61,0535	,5360	,9149
B5	38,4000	60,3673	,7003	,9112
B6	38,3400	60,9637	,6378	,9126
B7	38,2800	59,6751	,5716	,9143
B8	38,3200	61,2016	,5399	,9148
B9	38,4600	60,3351	,6690	,9117
B10	37,8000	62,4082	,3134	,9225
B11	38,4800	61,6016	,5501	,9145
B12	38,2200	60,6241	,6623	,9120
B13	38,1600	58,5861	,7496	,9093
B14	38,6600	63,2494	,4214	,9173
B15	38,4800	59,4384	,7048	,9106
B16	37,7400	61,8698	,4494	,9172
B17	38,5400	60,3759	,6231	,9127
B18	38,2400	59,4922	,7131	,9105

Reliability Coefficients

N of Cases =   50,0   N of Items = 18
Alpha =   ,9179

3.4.2 Interview Guidelines

As the main aim of this study was to get FAE instructors' perceptions of the ENG 101 Writing Exam Grading Criteria, a survey was carried out. As Brown & Rogers (2002) claim surveys typically take the form of interviews or questionnaires or both. This is why along with the questionnaire, the researcher carried out interviews. In other words, the aim here was triangulation since triangulation is something desirable in the research as viewing the same phenomena from multiple perspectives is possible in this way. (Brown & Rodgers,2002; Bailey, 1994; Cohen et al.,2000)

In this study, for the interviews, open-ended questions were prepared based on the items in the questionnaire. Later, a few more questions were added having analyzed roughly the common points that the instructors raised in the questionnaire. The aim of preparing the questions beforehand was to establish the reliability of the interviews as "One way of controlling reliability is to have a highly structured interview, with the same format and sequence of words and questions for each respondent" (Silverman, 1993; cited by Cohen et al 2000, p. 121).

The instructors who took part in the interview were volunteers. During the administration of the questionnaire, a piece of sheet was passed around and the instructors who volunteered filled in the chart on the paper by writing their full name, e-mail address and phone number so that the researcher could contact them. In total there were 12 instructors who volunteered but when they were called back, only 6 of them were able to arrange time for the

interview. The interviews lasted between 20 to 35 minutes. Interviews were held individually and tape-recorded for future reference.

3.5 Procedures

Before starting the research, the Directorate of Bilkent University was asked for permission to carry out this research. After permission was granted, the head of FAE program was consulted to get background information about the FAE program and the criteria itself. Since the criteria were recently introduced, the administration was also willing to receive such feedback before finalizing it. As a result, the questionnaire was prepared and it was approved by the head of FAE program. To raise the response rate of the questionnaire, all the instructors were given the questionnaire when they were all together for ENG 101 Course Evaluation session at the end of the semester. Before they started their meeting, the questionnaires were handed out to be filled in. However, as 9 of the instructors were teaching at summer school and had sessions at the same time, they couldn't attend the meeting and that's why only 55 out of 64 got the questionnaires and filled them in. As it was mentioned before, during the administration of the questionnaires, another piece of paper was passed around to be filled in by the instructors who would volunteer for the interviews.

After the questionnaire administration, the volunteers were called and appointments were made with each of them so that the researcher could get more detailed feedback about the ENG 101 writing exam grading criteria.

3.6 Data Analysis

Quantitative analysis was done for the first and the second section of the questionnaire using the Statistical Package for the Social Sciences (SPSS) program. As for the data for the open-response items in the questionnaire and the interview questions answers were subjected to content analysis and common themes was determined in the participants' responses (Miles & Huberman, 1994).

3.6.1 Analysis of the Questionnaire and the Interview Data

Data for this research gathered through the questionnaires which contained both closed- response items and open-response items. Apart from the questionnaires, interviews were carried out. As the questionnaire had both closed–response items and open-response items, the data analysis for the questionnaire was done both quantitatively and qualitatively. The first two sections of the questionnaire were analyzed statistically using SPSS data analysis program. For the first part of the questionnaire, descriptive statistics of biodata, frequency analysis and missing data analysis were done. Moreover, for the second part of the questionnaire, the reliability analyses were done. For the open-response questionnaire items and the interview data, descriptive categories, in other words headings, were developed from the data itself. To do this, all the responses for the questionnaires and the interviews on the sheets were transferred to the computer and under each heading recurring themes were noted down.

3.7 Limitations of the Study

First of all, since the main aim in this study is to get a certain group of instructors' perceptions on a certain criteria, the results cannot be generalized. This study is specific to the certain group which is FAE instructors at Bilkent University.

CHAPTER IV

DATA ANALYSIS AND INTERPRETATION OF RESULTS

4.0 Presentation

This chapter presents the analysis of the quantitative and qualitative data gathered from the questionnaires and the interviews, and ends up with the interpretation of the results.

4.1 Analysis of Quantitative Data

This section includes the data gathered from the second section of the questionnaire, where the instructors' opinions about the ENG 101 writing exam grading criteria were aimed to be found using a 1 to 4 likert scale. Since the main aim of this study was to find out the instructors' perceptions on the ENG 101 writing exam grading criteria in terms of its positive and negative attributes, during the data analysis, percentages of 'Strongly Agree' and 'Agree' sections were regarded as a whole and the same thing was applied for 'Disagree' and 'Strongly Disagree' sections. The data will be presented following the framework established in the research questions.

In the questionnaire, the first 8 questions were designed to get instructors' overall perceptions regarding the effectiveness of the criteria. Items nine and ten were specific about the categories and how the instructors view them. The next four items were mainly on descriptors. And the last four

items were written to find out participants' feelings about the criteria's application.

4.1.1 Instructors' Perceptions Regarding the Overall Effectiveness of the Criteria.

The first closed item in the questionnaire asked whether the common ENG 101 writing exam grading criteria are efficient for marking students' writing objectively across the FAE program. The aim here was to get instructors' overall views about the effectiveness of the criteria. As it is shown in Table 2, 80 % of the instructors believe that the criteria are good to assess students' writing objectively and fairly although the minority (20%) disagrees with the statement. In other words, the criteria seem to serve its purpose and this can be considered as a positive attribute of the criteria.

Table 2 Distribution of Responses to Item 1

The common *ENG 101 Writing Exam Grading Criteria* are good for marking students' writing *objectively* across the FAE program.	f	%
Strongly Agree	6	**12,0**
Agree	34	**68,0**
Disagree	8	**16,0**
Strongly Disagree	2	**4,0**
Total	50	**100,0**

The second item was about if instructors across the FAE program have a common understanding of the ENG 101 course writing objectives by having the common ENG 101 writing exam grading criteria. Sixty six percent of the instructors agree with this statement. However, 44 % of them do not agree with the idea that instructors across the program have a common

understanding of the writing objectives by having these criteria. This may mean that the criteria does not serve its purpose fully.

Table 3 Distribution of Responses to Item 2

By having the common *ENG 101 Writing Exam Grading Criteria* , instructors across the FAE program have *a common understanding of the ENG 101 course writing objectives.*	f	%
Strongly Agree	8	**16,0**
Agree	25	**50,0**
Disagree	11	**22,0**
Strongly Disagree	6	**12,0**
Total	50	**100,0**

The next question was aimed to find out if the criteria match with the ENG 101 course writing objectives or not. According to the results in Table 4, the majority of the instructors (76%) agree that the criteria reflect the writing objectives of ENG 101 course. This may also be regarded as something positive about the criteria.

Table 4 Distribution of Responses to Item 3

The common *ENG 101 Writing Exam Grading Criteria* reflect the *writing objectives to be covered in the ENG 101 course.*	f	%
Strongly Agree	9	**18,0**
Agree	29	**58,0**
Disagree	9	**18,0**
Strongly Disagree	3	**6,0**
Total	50	**100,0**

The fourth item in the questionnaire aimed to get the instructors' opinions about the practicality of the criteria. In Table 5, it is seen that 60 % of the instructors believe that the criteria help to save time while marking. Yet, a considerably high percent of the instructors (40%) do not agree with

this statement. This means that almost half of the instructors have difficulty in applying the criteria.

Table 5 **Distribution of Responses to Item 4**

Using the common *ENG 101 Writing Exam Grading Criteria saves time while marking.*	f	%
Strongly Agree	6	**12,0**
Agree	24	**48,0**
Disagree	19	**38,0**
Strongly Disagree	1	**2,0**
Total	50	**100,0**

The fifth item was to see if the criteria are user- friendly or not. In Table 6, it is shown that the great majority (76%) agree that the criteria are user-friendly. That means the majority of them can use the criteria easily.

Table 6 **Distribution of Responses to Item 5**

The common *ENG 101 Writing Exam Grading Criteria* are *user-friendly.*	f	%
Strongly Agree	5	**10,0**
Agree	33	**66,0**
Disagree	11	**22,0**
Strongly Disagree	1	**2,0**
Total	50	**100,0**

The sixth item aimed to get feedback from the instructors about whether the criteria successfully discriminate border- line pass and border-line fail papers. According to the Table 7, 72 percent of the instructors state that the criteria are good at discriminating border- line pass and fail papers. This may also be taken as a positive attribute of the criteria.

Table 7 Distribution of Responses to Item 6

The common *ENG 101 Writing Exam Grading Criteria* are good at *discriminating between border-line pass and border-line fail papers.*	f	%
Strongly Agree	4	8,0
Agree	32	64,0
Disagree	13	26,0
Strongly Disagree	1	2,0
Total	50	100,0

The next question aimed to find out if having the common criteria enhances fair marking of students' papers across the FAE program or not. As mentioned in the first chapter, FAE has five different units and each had their own criteria and that led to some problems across the units. However, the participants' responses show that the majority of the instructors (72 %) believe that the criteria enhance fair marking across the program (see Table 8). This result shows that the criteria help the instructors to decide on the pass and fail papers.

Table 8 Distribution of Responses to Item 7

Having the common *ENG 101 Writing Exam Grading Criteria* enhances *fair marking of students' papers across the FAE program.*	f	%
Strongly Agree	9	38,0
Agree	20	40,0
Disagree	19	18,0
Strongly Disagree	2	4,0
Total	50	100,0

Item 8 is one of the 'reliability–check' items in the questionnaire since it paraphrase the sixth item. Similar to the sixth item's percentage, 70 % of the instructors believe that using these criteria they can decide easily for the papers which pass or fail.

Table 9 Distribution of Responses to Item 8

Using the common *ENG 101 Writing Exam Grading Criteria* instructors can *decide easily* whether a student's paper passes or fails.	f	%
Strongly Agree	5	10,0
Agree	30	60,0
Disagree	13	26,0
Strongly Disagree	2	4,0
Total	50	100,0

4.1.2 Instructors' Perceptions Regarding the Categories

Items 9 and 10 are about categories in the criteria. As seen in Table 10, a high percentage of the instructors (78%) agree that the categories match with ENG 101 Course writing objectives. However, 22 % of them still think that they do not match with the writing objectives. This may also regarded as one of the positive attributes of the criteria.

Table 10 Distribution of Responses to Item 9

In the criteria, *the categories-* content- organization- language and citation/ tech- *match the ENG 101 course writing objectives.*	f	%
Strongly Agree	7	14,0
Agree	32	64,0
Disagree	10	20,0
Strongly Disagree	1	2,0
Total	50	100,0

In the next item, the equal weighting of the categories are questioned and the instructors' perceptions were asked. Sixty four percent of the instructors disagree with the idea that it is good to have all categories-content- organization- language- and citation / tech. equally weighted. This is one of the major issues which was also raised in the open-response sections. This is the most negative attribute of the criteria as perceived by the participants.

34

Table 11 Distribution of Responses to Item 10

It is good that all categories- content- organization- language and citation/ tech- in the criteria *have equal weight* while assessing students' written products.	f	%
Strongly Agree	4	**8,0**
Agree	14	**28,0**
Disagree	22	**44,0**
Strongly Disagree	10	**20,0**
Total	50	**100,0**

4.1.3 Instructors' Perceptions Regarding the Descriptors

Items 11, 12, and, 13 were all about the descriptors in the categories. According to Table 12, a vast majority (84%) of the instructors believe that the descriptors match the ENG 101 course writing objectives. This may also be stated as another positive attribute of the criteria.

Table 12 Distribution of Responses to Item 11

The descriptors (the bullet points for each category in each band) *reflect* the ENG 101 course writing *objectives.*	f	%
Strongly Agree	6	**12,0**
Agree	36	**72,0**
Disagree	6	**12,0**
Strongly Disagree	2	**4,0**
Total	50	**100,0**

Item 12 asked if the descriptors discriminate well between bands and within each band or not. In Table 13, it is shown that only 58 % of the instructors agree with this statement. On the other hand, a considerably high number of the instructors disagree with the statement. In other words, there may be a problem with the descriptors.

Table 13 Distribution of Responses to Item 12

The *descriptors* discriminate well between bands and within each band.	f	%
Strongly Agree	4	8,0
Agree	25	50,0
Disagree	21	42,0
Total	50	100,0

The next item aimed to find out if the descriptors help instructors assign grades easily or not. Although 58 % of the instructors believe that the descriptors discriminate well between the bands, a considerably high percent (42%) of them do not agree with the others. Based on Table 14, we may conclude that there may be problems with the descriptors for some instructors.

Table 14 Distribution of Responses to Item 13

The *descriptors* under each category in each band let me assign the appropriate grades *easily* to students' papers.	f	%
Strongly Agree	4	8,0
Agree	25	50,0
Disagree	18	36,0
Strongly Disagree	3	6,0
Total	50	100,0

Item 14 is about the wording of the descriptors. In Table 15, a large number of the instructors (88%) agree with the statement that the wording is easy to understand. This may mean that the instructors do not have difficulty in understanding the descriptors. This may also be considered as something positive about the criteria.

Table 15 Distribution of Responses to Item 14

The *wording in the descriptors* is *easy* to understand.	f	%
Strongly Agree	11	**22,0**
Agree	33	**66,0**
Disagree	6	**12,0**
Total	50	**100,0**

4.1.4 Instructors' Feelings about the Criteria's Application

In the questionnaire, the last four items aimed to get the general feelings of the instructors' about the application of the common ENG 101 writing exam grading criteria.

As it is seen in Table 16 and Table 18, a vast majority (80%) of the instructors feel confident about using the common criteria appropriately.

However, looking at Table 17, it can be concluded that most of them (68 %) do not feel that all the other raters / instructors across the FAE program have a common understanding of the criteria. This may mean that although the criteria aim to enhance standardization across the FAE program, it has not achieved its purpose yet.

Table 16 Distribution of Responses to Item 15

I am *confident* about using the common ENG *101 Writing Exam Grading Criteria* appropriately.	f	%
Strongly Agree	8	**16,0**
Agree	32	**64,0**
Disagree	8	**16,0**
Strongly Disagree	2	**4,0**
Total	50	**100,0**

Table 17 Distribution of Responses to Item 16

I am *confident* that *all of the raters/ instructors across the program* have a common understanding of the common *ENG 101 Writing Exam Grading Criteria*	f	%
Strongly Agree	1	**2,0**
Agree	15	**30,0**
Disagree	26	**52,0**
Strongly Disagree	8	**16,0**
Total	50	**100,0**

Table 18 Distribution of Responses to Item 17

I am confident that I have a clear understanding of the common *ENG 101 Writing Exam Grading Criteria.*	f	%
Strongly Agree	10	**20,0**
Agree	30	**60,0**
Disagree	9	**18,0**
Strongly Disagree	1	**2,0**
Total	50	**100,0**

Finally, in Table 19 it is seen that although most of the instructors (68%) seem to be satisfied with the common ENG 101 writing exam criteria, 32% of the instructors are not for various reasons.

Table 19 Distribution of Responses to Item 18

I am satisfied with the common *ENG 101 Writing Exam Grading Criteria*	f	%t
Strongly Agree	3	6,0
Agree	31	62,0
Disagree	13	26,0
Strongly Disagree	3	6,0
Total	50	100,0

To find out why a considerable number of instructors at some points disagreed with the statements open –response items were included in   the questionnaire. As well as the open- response items some interviews were

carried out to have more detailed feedback on the criteria. In the following section a qualitative analysis of these data will be presented.

4.2 Analysis of the Qualitative Data

The qualitative data in the questionnaire came from the third section (C), in which a scenario was given to find out if the instructors apply the criteria in the same way while marking or not, from the fourth section (D), where the instructors were asked to write down in general the strengths and points to consider about the criteria, and from the last section (E) where the instructors were supposed to note down the suggestions they might have to improve the common ENG 101 writing exam grading criteria.

Apart from the data gathered from the questionnaire, qualitative data were gathered through the interviews as well. In the study, while discussing the results, the original quotations taken from the questionnaire were labeled as *respondent x* and the interviews as interviewee x. However, this does not mean *interviewee x* and *respondent x* are the same people.

4.2.1 Analysis of Responses to Section C in the Questionnaire

Section C in the questionnaire was designed to answer the second research question of the study. Out of 50 valid questionnaires, only in 40 of them this section was filled in. All the responses were transcribed as they were written originally in the questionnaire without making any corrections. Then, recurring themes and topics were identified through a process of

coding and categorizing to answer how the instructors apply the same common ENG 101 writing exam criteria.

Parallel to the results obtained in the 16th item where most of the instructors did not feel confident that all raters/ instructors across the program have a common understanding of the criteria, some instructors also accepted and restated their opinion that the case in the scenario was something important to be addressed so that all instructors across the FAE program have a common understanding of the criteria and its application.

".... I think this is a serious issue." (Respondent 15)

"The scenario you point out is a serious issue to be considered."
(Respondent 39)

"I believe this scenario needs to be discussed thoroughly within the units and FAE to have a common understanding." (Respondent 53)

Table 20 Distribution of the Responses to Section C

The scenario:
Suppose that one student's paper matches the B band in two categories* but merits a C-band in the other Two. How would you mark the paper using the common *ENG 101 Writing Exam Grading Criteria* and what grade would you give to that paper and why?
 	
I would mark the paper by…
   • Depending on how successful the student is at ***content and organization (19)***
   • Depending on how successful the student is at ***content and language (7)***
   • Taking the average arithmetically. **(14)**

When the recurring answers were analyzed, it is seen that a vast majority of the instructors consider the weighting of the categories in a different way and in one way or another they take an average to give a final grade to the paper.

When the results were analyzed in detail, it was seen that almost 48% of the instructors pay attention to content and organization more while marking:

"It would depend on which categories were better. Content and organization are more important so if those categories are better in the B range I would give a low B. If the language and citation / tech. categories are B but there are serious problems with the content and organization, the student would get a C." (Respondent 54)

"I would consider content and organization as priorities and reconsider the paper. I think these two categories represent the basic conventions that need to be implemented in an essay…" (Respondent 53)

Although almost 48 % of the instructors pay attention to *content* and organization, 18 % of the instructors believe that apart from *content,* not *organization* but *language* is important:

"I would give it a B providing the language was clear enough to express ideas and argument intelligibly."(Respondent 8)

"*It would depend on which categories the student had performed at a C- level. If it was content and language I think I would be inclined to give a grade somewhere around the bottom end of C level. This is why some weighting may need to be introduced." (Respondent 39)*

"I would focus on content and language- if it was high in these two bands, I would use this to average the grades."(Respondent 17)

Despite the fact that instructors give importance to different categories, 25% of the instructors preferred to take the average in order to be fair to the students:

"I would grade content, organization, language, citation/ tech. separately and take their average for the grade of the paper." (Respondent 23)

"This of course depends on individual papers but I would give average of four bands." (Respondent 28)

"To be on the safe side, I would convert each grade to its numerical equivalent. Calculate and divide by 4, then, re-convert to letter grade." (Respondent 31)

As it can be understood from the numbers and quotations from the questionnaires, there is no common way to use the criteria. Although the aim of the new criteria was to have a common understanding and marking across the FAE program it seems that the aim has not been achieved completely yet. In other words, there are still discrepancies in the way the instructors apply the criteria.

4.2.2 Analysis of the Data from the Comments in Section D

When the questionnaire was designed, the aim of section D was to find answers for the third and fourth research questions. In other words, positive and negative attributes of the common Eng 101 writing exam grading criteria were the main interest. When the data were considered, it was found that not every participant filled this section in. Only 28 of them noted down some issues as strengths of the criteria and the points to consider about the criteria. The strengths of the criteria were taken as the positive attributes of

the criteria whereas points to be considered were regarded as the negative attributes of the questionnaire during the analysis.

In the analysis of section D, the answers were grouped under the following headings in two aspects as positive and negative.

- Overall comments
- Categories
- Bands
- Descriptors
- Match between Eng 101 course writing objectives across the FAE program and the criteria.

In Table 21, the distribution of the responses is shown. The numbers in parenthesis show total number of participants who mentioned each issue.

Table 21 Distribution of the Responses to Section D

	POSITIVE	NEGATIVE
OVERALL FEEDBACK FOR THE ENG 101 WRITING EXAM GRADING CRITERIA	• Effective guide for both students and teachers as it sets out the most important points to consider. (4) • Good for standardization across the FAE program.(12) • Simple not complicated.(7) • Saves time. (3) • Should be a model for other tasks / assignments.	• Should be more specific and less open to interpretation more simplified. • More detailed criteria would be better.
CATEGORIES IN THE ENG 101 WRITING EXAM GRADING CRITERIA		• They shouldn't have equal weighting. (14) • More emphasis on thesis, topic sentence, development ideas, transitions, conclusion. • More points should be allocated for content and organization. (12) • The breakdown of each category needs to be revised. The content, organization and language parts should be given a higher percentage in 101. (8)
BANDS IN THE ENG 101 WRITING EXAM GRADING CRITERIA	• Good description and various level of proficiency. (3) • Enables us to discriminate the borderline pass and fail papers. • Good to have C- defined.	• I am not satisfied with F band. (5) • More discrimination within the F band. • Grade by numbers not letters.
DESCRIPTORS IN THE ENG 101 WRITING EXAM GRADING CRITERIA	• They are satisfactory/ ok. (11) • Good in general –still be fine tuned. • Clear descriptors to fairly evaluate students' products.	• Sometimes, the difference among the descriptors is notably slight. • C pass is not clear, open to interpretation. • Some points could be added. • Somehow open to interpretations. • The descriptors between categories can be more precise.
MATCH BETWEEN THE ENG 101 COURSE OBJECTIVES AND THE DESCRIPTORS IN THE CRITERIA	• Reflects the objectives.(13)	• There should be more emphasis on introduction, paragraphing and transitions

In general, instructors seem to be quite satisfied with the existence of such common criteria across the FAE program. They believe that the criteria help both the instructors and the students as two of the instructors wrote down in the questionnaire:

"Potentially effective guide for both students and instructors." (Respondent 18)

"It guides instructors and students; it eliminates confusion." (Respondent 23)

"It sets out the most important points to be considered." (Respondent 2)

Most of the instructors also believe the criteria help for standardization across the program:

"It is useful to achieve standardization within the FAE program".

(Respondent 43)

"The criteria set a common, easily identifiable standard." (Respondent 40)

"It is good that all students are marked according to the same criteria."

(Respondent 34)

Although very few of them think just the opposite, most of the instructors find the criteria user-friendly and simple to understand:

" It is not so wordy and complicated". (Respondent 18)

" allows quick decision to be made".(Respondent 11)

However, looking at the results of both items nine and ten in section B and the points mentioned in section D, it is clear that a majority of the

instructors do not agree with the idea that all categories should have equal weighting:

> *"They shouldn't have equal weighting. I believe 30   30   30 10 breakdown would be more consistent with FAE objectives". (Respondent 12)*

> *"More points should be allocated for content, organization and language."*
>
> *(Respondent 18)*

> *"The breakdown of each category needs to be revised. The content, organization and language parts should be given a higher percentage in 101." (Respondent 2)*

> *"More emphasis on content and organization is needed." (Respondent 45)*

Parallel to the findings in section B, some want to have more emphasis on content and organization whereas some want to have more emphasis on content and language. In short, it may be concluded that the instructors are not happy with the weighting of the categories.

In terms of bands, the criteria seem to be satisfying:

> *"easy to distinguish A/ B B/C D /F."(Respondent 28)*

> *"Good description of various levels of proficiency."(Respondent 55)*

> *"Enables us to discriminate borderline pass and fail papers." (Respondent 8)*

Although most of the instructors seem to be satisfied with the bands, they are not happy with F band. They look for more discrimination within F band. Most of them believe that 54 is too high to get an F.

> *"I am not satisfied with F band."(Respondent 41)*

"More discrimination is needed in the F band." (Respondent 33)

Related to the bands, most instructors state that marking by letters is not practical since they grade by numbers while entering the grades onto the computers.

"Our final result needs to be a percentage, so grade bands are artificial."

(Respondent 48)

"Letters in the bands do not make sense as we focus on numbers to make the final calculations."(Respondent 9)

With regard to descriptors, the instructors seem to be satisfied. They believe that they are clear and easy to understand:

"Clear descriptors to fairly evaluate students' products". (Respondent 7)

"Descriptors are satisfactory."(Respondent 54)

"Descriptors are good in general."(Respondent 32)

On the other hand, there are some instructors who believe that the difference among descriptors is notably slight in some. They also add the descriptors are somehow open to interpretations.

"Some descriptors still need to be fine tuned."(Respondent 20)

"In the criteria descriptors are somehow open to interpretations." (Respondent 6)

"The descriptors between bands can be more precise."(Respondent 48)

Finally, when the match between ENG 101 course objectives and the descriptors in the criteria are taken into consideration, it may be seen that almost all instructors agree that the criteria reflect ENG 101 course objectives.

4.2.3 Participants' Suggestions to Improve the Criteria Based on the Data from the Comments in Section E

Section E in the questionnaire aimed to answer the fifth research question which aimed to obtain participants' suggestions to improve ENG 101 writing exam grading criteria. However, a vast majority of the participants (76%) did not write any suggestions to improve the criteria. In other words, only 12 of them expressed their ideas on how to improve the criteria. However, this is not surprising because when we refer to the results of the other sections i.e. first 6 questions in section B or positive attributes of the criteria concluded by the results of section D, in general instructors are happy with the criteria though some believe it needs some slight changes (e.g. weighting of the categories).

As a suggestion, one of the instructors wrote that 'plus' and 'minus' zones should be added. Some others suggest that the weighting in the categories needs to be reorganized. Another instructor would like to have more discrimination within the F band.

More important than that, out of twelve eight instructors called the attention to the training that the instructors need on how to use these criteria:

"The criteria could be more specific and that can be the first step to train the instructors on how to use the criteria. For two semesters now we are using the same criteria and sometimes there are enormous gaps / differences between instructors' assessment". (Respondent 41)

Another issue raised as a suggestion was that grading needs to be by numbers – not by grade bands. Five instructors stated that final results needed to be a percentage, so grade bands were artificial. As a result, they would like to have more analytical   criteria.

Apart from some suggested slight changes in the criteria, the instructors want a standard way to mark the papers.

"The management of the grading appears to change too often. For example, for once, only the instructors grade their students but for the next time, the instructors never grade their own students. For another time instructors grade their student together with another instructor etc. The management needs to be finalized so the program can settle down". (Respondent 28)

In other words, it may be concluded that there needs to be some ground rules on how to carry out the marking as well.

4.2.4 Analysis of the Interview Data

In addition to the questionnaire, interviews were held with some instructors for mainly 'methodological triangulation' (Brown & Rogers, 2002, p.244) and to be able to get more detailed data. Although 12 instructors volunteered to take part in the interviews, half of them were not available when they were asked to have the interviews. As a result, only six of them were interviewed.

The interview was composed of 12 questions aimed at getting more data about the instructors' perceptions on the criteria. All the interviews were audio-taped for future reference. The results of the interviews will be presented in relation to the 12 interview questions asked. Since the majority of the interviewees gave similar responses, their ideas were grouped and some sample ones representing   the group will be presented here.

The first question was asked to find out whether the instructors were aware of the rationale of introducing 'the Common ENG 101 writing exam grading criteria'. All six interviewees gave the similar answer as this was due to having a standard among five different units in the FAE.

Some of the responses were as follows:

"I believe …to make an attempt at standardization across FAE for the final exam."

(Interviewee 2)

"I suspect it is in order to ensure that the students within units and across units are not negatively affected by extremely difficult marking."

(Interviewee 3)

" For standardization of course......" (Interviewee 1)

"To create common understanding. Especially if you are new to the program like me, this is really beneficial."(Interviewee 4)

Some of the instructors also added that they had some problems since the students compare the marking done in different unit and by different instructors. They were now happy to have standard expectations from the students across the units.

The second question aimed to find out if the instructors were for or against the idea of having a common writing exam grading criteria. Most of them agreed with this idea:

"I think it is a good idea for the common exam. I would not agree with that for every assignment but for the common exam… yes…good idea." (Interviewee 3)

"It is good for both teachers and students. Teachers know better what to focus on and students know what is expected from them regardless their departments."

(Interviewee 6)

The third and fourth questions were about descriptors. In general the descriptors were satisfying for all the instructors. However, very few of them mentioned that a couple of the descriptors need to be rephrased:

"The descriptors are done fairly well. Descriptors are continuous throughout the bands…However, we should not forget that there are no perfect criteria but as a foundation this is good." (Interviewee 4)

"They are simple and easy to interpret. However, there are some overlaps. i.e. in band A it says 'answers the question set clearly and thoughtfully' and in B band it says 'answers the question set clearly and in detail.' " (Interviewee 1)

For the next question which focuses on the match between the ENG 101 course writing objectives and the criteria, all instructors share the idea that the criteria match ENG 101 writing objectives to a great extent:

"Yes…they reflect the objectives quite well…" (Interviewee 5)

"I think they generally match…However, it has to employ the academic wording we use in our classes. E.g. thesis statement and topic sentences…."

(Interviewee 1)

"This reflects most of the objectives. It cannot reflect all…" (Interviewee 4)

In question six, when the instructors were asked to what extent the criteria reflect actual teaching done in class, they claimed that since in class their aim is to cover the objectives, it does reflect the actual teaching done in class:

"It fairly does."(Interviewee 1)

"Reasonably well…   "(Interviewee 2)

"to a large extent… sure… it does…for example we focus on how to write a good introduction or thesis statement or how to cite academically…." (Interviewee 6)

The seventh and eight questions were added after analyzing the results of the questionnaire roughly. When the results of item 10 were considered in the questionnaire it is seen that most of the instructors are not happy with the equal weighting among the categories. The instructors in the interview also confirmed this finding:

"I do not like equal weighting of the categories. " (Interviewee 2)

"I think they should have different weighting according to the focus determined by the task "(Interviewee 1).

"I believe giving information and locating information is important because we are giving content-based instruction. To me, content should be out of 40, organization 30, language 20 and citation 10." (Interviewee 5)

In the eighth question, when the instructors were asked if there were any categories that they paid more or less attention while marking, they confessed that they all focused more on content and organization:

" content and organization are more important I believe… You can always fix the language or punctuation or etc. …" (*Interviewee 4*)

The next interview question asked whether they feel confident that there is a common understanding of the criteria among all the raters / instructors across the program or not. Unfortunately, none of them feel confident about this issue:

"I think no…" (Interviewee 3)

" No, I do not.." (Interviewee 1)

This result was not surprising for the researcher as similar results were found in the analysis of Section B items 16, 17 and 18. The instructors also add that the reason why they do not feel confident is because they observed serious gaps between the grades of the instructors who are cross-grading after the final exam. In the same question, the instructors were also asked to come up with some ideas to overcome such a problem:

" I think the only solution when there is discrepancy a third person should sit and mark the paper." (Interviewee 3)

" Training should be provided in a more organized and controlled way. "

(Interviewee 1)

Since these criteria were launched not a long time ago, the 10th question was asked to see if the instructors were given any training on how to use the criteria. It was interesting to note that although all of them agreed that they had some sort of training, they were not satisfied with it:

"we had a meeting… and…. we were given some sample papers to mark…. we spent almost three hours but we did not walk out standardized"(Interviewee 6)

Question 11 aimed to get instructors' perceptions on the strengths of the criteria. They were all positive despite some of the weaknesses of the criteria that they mentioned in the last question. Some of their responses about the strengths of the criteria were as follows:

"Its existence is a strength- it does give people some idea about what they are aiming for…" (Interviewee 4)

"It reflects the objectives.."(Interviewee 2)

"It is fair and helps objective marking across the program…"(Interviewee 5)

"Teachers feel confident on what to focus and students feel confident as they know what is expected from them." (Interviewee 6)

"It is user-friendly and can be applied practically." (Interviewee 1)

In terms of the weaknesses of the criteria, the instructors agree that some minor changes need to be done in the descriptors but more important than that the weighting of the categories need to be revised:

"I believe the most important thing is the weighting of the categories… I mean … they need to be reorganized."(Interviewee 4)

Another issue that the instructors raised in the interview but not in the questionnaire was about letter grades. Since the calculation is by numbers and they use numbers when they enter the grades onto the computers, they do not want to have letter grades:

"…We should get rid of letter grades. They are confusing…or minus and plus bands need to be added since the range between grades in each band is huge…"(Interviewee 2)

"..F band is problematic. The difference between 54 and 0 can make a huge difference to a student's final grade…" (Interviewee 5)

4.3. Comparison of Questionnaire and Interview Results

In this study, two data collection techniques -a questionnaire and interview- were used to find out the instructors' perceptions of the ENG 101 writing exam grading criteria. The aim of using two different techniques was to have methodological triangulation (Brown & Rogers, 2002). When the data from the questionnaires were compared with the data from the interviews, it was seen that they were consistent and parallel to each other. In both, it was found that in general the instructors were happy with the criteria and they were all aware of the rationale behind having common criteria for ENG 101 course writing exam. In both, they stressed the standardization and having a common understanding across the program. In terms of the categories in the questionnaire and in the interviews they stated that they really did not like the

idea of having all the categories equally weighted. This may be regarded as the major point to be considered about the criteria. In other words, this may be seen as one of the negative attributes of the criteria. Although the results in both seem to be parallel, there was an interesting point about the descriptors. In the questionnaire, 88% of the instructors stated that the descriptors were easy to understand. However, in the interviews almost all of them stated that the some descriptors were confusing and they need to be revised. This may be because when they were filling in the questionnaires, they just roughly expressed their perception of the descriptors. On the other hand, during the interviews they had more time to look at the descriptors in detail and so they could tell more about the quality of the descriptors. Finally, when the participants' feelings about the criteria were considered, the results match to a great extent. In both, they stated that they themselves feel confident about using the criteria appropriately but they are not sure about their colleagues since they observed gaps among themselves while marking.

CHAPTER V

CONCLUSION

5.0 Presentation

This chapter presents the summary of the study, discussion of the findings, implications and recommendations for further studies.

5.1 Summary of the Study

The main aim of the study was to find out the instructors' perceptions on the ENG 101 writing exam grading criteria, which was designed recently to be used across the FAE program. Mainly two instruments were used; a questionnaire having 5 sections each addressing a different research question and the interviews held with the instructors on voluntary base. First, the questionnaire was designed in the light of the key points given in the literature about how to carry out research and develop instruments (Brown & Rodgers, 2002; Bailey, 1994; Cohen et al 2000; Patton,1990; Ritchie and Lewis,2004; Brown, 1988). Then, the questionnaires were administered to 55 FAE instructors at the same time to increase the response rate. Meanwhile, they were asked if they would volunteer to take part in the interviews. The ones who volunteered filled in a sheet where they wrote their full name, e-mail address and phone numbers so that the researcher could get into contact with them. Next, the interview guidelines were prepared in the light of research questions and the results that came out after going over the

questionnaires thoroughly. Twelve questions were asked in the interviews. They were all parallel to the research questions and the aim was to get more detailed answers. After gathering the data the results were analyzed both quantitatively-using SPSS- and qualitatively.

5.2 Discussion of the Findings

The first research question in this study was how Eng 101 instructors perceive the common Eng 101 writing exam grading criteria in terms of the following dimensions;

- Overall effectiveness

- Categories

- Descriptors

- Participants' feelings about its application

Regarding all the findings from the questionnaires and the interviews, it may be concluded that most instructors were generally satisfied with the new criteria. They believe that the criteria help to have standard grading across the FAE program. However, they still have some doubts about the way that the criteria are applied across the program while assessing students' papers. This fact cannot be denied as some instructors claimed that they have a different approach while using the criteria in their units:

> *"We grade differently in our unit. Weighting goes from left to right in terms of priority, so it depends. If we consider that they are equal, it should be C+ or C."*
>
> *(Respondent 16)*

"Because not all bands / categories have equal weight in our unit it would depend on which areas were higher or lower. Also there are specific penalties for such errors as plagiarized passages, no works cited pages etc." (Respondent 28)

Since this is the case in one or more units, this is a serious issue to be resolved. This fact totally contradicts the aim of having such common criteria across the units in the FAE program.

In other words, when findings to the second research question which was about the way the instructors mark the papers are taken into consideration, there are quite different approaches adopted. Although being satisfied with the criteria in general in terms of overall effectiveness, descriptors bands and etc. which can be regarded as positive attributes of the criteria, the difference among instructors in the way they apply the criteria can be considered as a negative attribute of the criteria. These findings match with Lumley's (2002) conclusions after his study to find out what assessment criteria really mean to the raters:

…although there appears to be some evidence that the raters understand the rating category contents similarly in general terms, there is also evidence that they sometimes apply the contents of the scale in quite different ways. They appear to differ in the emphasis they give to the various components of the scale descriptors…(p.266)

In terms of other dimensions there seem no problems except the equal weighting of the categories and the instructors' feelings about the application of the criteria.

Reflecting on the data gathered from the last section of the questionnaire and the answers to the interview questions, there were a couple of suggestions that the instructors came up with. The most significant of these was about training and this is also another fact that matches with what Lumley (2002) suggests. According to him, training plays an important role in influencing raters' behaviors, especially by clarifying rating criteria. When the ENG 101 instructors' suggestions for training and standardization session are taken into consideration, the literature also supports this idea. Weigle (1994; cited in Lumley, 2002) found that rater reliability increased as a result of training and that improved agreement was the result of raters gaining better consensual understanding of the terms and levels represented in the scale.

Apart from the training or standardization sessions they look for, most of the instructors do not want the categories in equal weighting. The results show that almost all of them believe that 'content' has the priority to be achieved by the students. This means 'content' requires higher grade or percentage in the criteria. To have a common understanding, this problem needs to be solved explicitly or else the instructors use their intrinsic criteria. Another problem which can be regarded as a negative attribute of the criteria is about the letters used for bands.. They believe this is confusing for them as they make the final calculations in numbers so having letters sounds meaningless for them.

5.3 Implications and Recommendations

Having the data in hand regarding the participants' perceptions and suggestions for the criteria, there are two main recommendations that could be offered. Firstly, the categories may have different weightings. Based on the results, it is suggested that 'content' may have the highest weighting and followed by 'organization' and 'language' respectively. A second suggestion is that slight changes need to be made in the wording of the descriptors. Some of the adjectives used are quite similar to each other as one of the instructors in the interview exemplifies:

> *"Although descriptors are ok in general very few need to be reworded. For example, here, 'powerfully' and here 'thoughtfully'….The distinction needs to be made clearer…"(Interviewee 1)*

Finally, since the main and most important aim to design these criteria was to have a common understanding of the writing objectives across the program and assessing students in the same way with set criteria, there is a need for training the instructors on how to apply the criteria. Almost all the instructors support this idea. Although they had already been given training once, they believe that it was not effective.

> *"Instead of giving training to huge groups of instructors altogether in a hall, in small groups as many sample papers as possible should be marked so that we can come to an agreement. Also, the ground rules for the criteria should be set by the trainers but should not be open to discussion." (Interviewee 1)*

Thus, in the FAE properly designed standardization sessions could be conducted in order to ensure that "raters use the scale appropriately and

consistently" (Weigle, 2002, 108). For standardization, first, the leader or preferably a team should read through the scripts to find anchor/ benchmark scripts that exemplify the different points on the criteria. In this context, the head of FAE and the heads of the units could come together and decide on the anchor scripts. It would also be helpful to include in the training sets scripts that exemplify certain problematic situations, for example, scripts that do not respond to the task or simply copy the prompt, or scripts that represent the borderline between two critical levels such as pass and fail. It would be important that anchor papers illustrate the nuances of the criteria. Next, other instructors may be asked to use the criteria and the anchor papers to evaluate a sample set of responses. Any discrepancies between the scores that are assigned by the instructors should be discussed. The discussions could be done in groups. However, it should be noted that it is virtually impossible to get a large group of raters to agree on exact scores and that some disagreement is inevitable. As well as in groups, the raters may also be asked individually to justify why they assign that score to the script. Last but not least, raters who consistently rate higher or lower than the rest of the group should be given feedback and perhaps additional training to bring their scores into alignment with the rest of the group (Weigle, 2000; Moskal &Jon, 2000).

In conclusion, the ENG 101 writing exam criteria has both positive and negative attributes as mentioned throughout the study. Considering the

findings as a result of this study, the relevant people can make the necessary changes to improve the criteria and use it more efficiently.

Some recommendations can also be made if similar research is to be carried out. Although the results were quite satisfying and motivating for these recently launched criteria, more accurate feedback could have been obtained if the instructors were asked to mark same papers under the same conditions and the grades could be compared and discussed. Due to some constraints such as time and human resource, such a study could not be added to support the idea that the ENG 101 writing exam criteria is reliable. Last but not least, in terms of instruments used 'Think- aloud protocol' could have been applied as it may allow analysis of such things as the sequence of rating, the interpretations the participants make of the scoring categories in the criteria and the difficulties raters face in rating etc.

5. 4 Further Research

In this study, the focus was the FAE instructors' perceptions of the ENG 101 writing exam grading criteria. More specifically, positive and negative attributes of the criteria were explored. However, a new research study can be conducted to find the reliability of the criteria. "The two forms of reliability that are typically considered in classroom assessment and in scoring rubric development involve rater (or scorer) reliability. They are interrater and intrarater reliability" (Moskal & Jon, 2000, p.7). So, the interrater reliability

and intrarater reliability of the ENG 101 writing exam grading criteria could be

studied in another research.

REFERENCES

Alderson, J. C. & Clapman, C. & Wall, D. (1995). *Language test construction and evaluation.* Cambridge: CUP.

Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice.* Oxford: Oxford University Press.

Bailey, K. D. (1994). *Methods of social research.* New York: The Free Press.

Brown, J. D. (1988). *Understanding research in second language learning.* Cambridge: Cambridge University Press.

Brown, J. D. (1996). *Testing in language programs.* Upper Saddle River: NJ: Prentice Hall Regents.

Brown, J. D. & Rodgers, S. T. (2002). *Doing second language research.* Oxford: Oxford University Press.

BUSEL (2004). *ENG 101 course objectives.* Retrieved from the web 1/ 12/ 2004 http://www.bilkent.edu.tr/~busel/faep.htm

Cohen, L., Manion L. & Morrison K. (2000). *Research methods in education.* New York: Routledge Press.

Crusan, D. (2002). An assessment of ESL writing placement assessment. *Assesing Writing,* 8, 17-30.

Cumming, A. (1990). Expertise in evaluating second language compositions. *Language testing* 7*, 31-51.*

Douglas, D. (2001). Language for specific purposes assessment criteria: where do they come from? *Language Testing,* 18,171-185.

Fraenkel, J. R. & Wallen, N. E. (2000). *How to design and evaluate research in education.* New York: McGrawHill.

Hawkey, R. & Barker, F. (2004). Developing a common scale fort he assessment of writing. *Assessing Writing,* 9,152-159.

Heaton, J. B. (1988). *Writing English language tests.* New York: Longman Group UK limited.

Hughes, A. (1989). *Testing for language teachers.* Cambridge: CUP.

Lumley, T. (2002). Assessment Criteria in a large-scale writing test: what do they really mean to the raters? *Language Testing,* 19, 3, 246-276.

Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis.* SAGE, Thousand Oaks.

Moskal, B. M. & Jon, A. L. (2000) Scoring Rubric Development: Validity and Reliability. Practical Assessment, Research & Evaluation, 7 (10). Retrieved from the web 6/9/2005 http://PAREonline.net/getvn.asp?v=7&n=10

Oppenhaim, A. N. (1992). *Questionnaire design, interviewing and attitude measurement.* New York: Pinter Publishers.

Park, T. (2004). *Scoring procedures for assessing writing.* Retrieved from the web 5/12/2004. http://www.tc.columbia.edu/academic/tesol/webjournal/park_Forum.pdf

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. UK: SAGE Publications Ltd.

Perkins, K. (1983). On the use of composition scoring techniques, objective measures, and objective tests to evaluate ESL writing ability. *TESOL Quarterly*, 17, 651-666.

Ritchie, J. & Lewis, J. (2004). *Qualitative research practice: A guide for social science students and researchers*. UK: SAGE Publications Ltd.

Turner, C. E. (2000). Listening to the voices of rating scale developers: Identifying salient features for second language performance assessment. *The Canadian Modern Language Review,* 56, 555-584.

Weigle, S. C. (2002). *Assessing writing*. Cambridge: CUP.

Wharton, S. (2003). Defining a appropriate criteria for the assessment of master's level TESOL assignment. *Assessment and Evaluation in Higher Education,* 28,649-663.

White, E. M. (1984). Holisticism. *College Composition and Communication*,35, 400-409.

White, E. M. (1994). *Teaching and assessing writing*. San Fransisco: Jossey-Bass Publishers.

Wier, C. J. (1993). *Understanding and developing language tests*. New York: Prentice Hall.

William, D. J. (1996). *Assessing writing. Preparing to teaching writing*. Belmond: Lawrence Erlbaum Associates.

APPENDICES

APPENDIX A

ENG 101 WRITING OBJECTIVES

Academic Writing

By the end of Eng 101, students should be able to: (pre-writing)

• analyse a given title / task / question fully, carefully and in sufficient detail.

• consider specific information and counter viewpoints in detail before planning an essay

• decide on a suitable viewpoint / focus for their writing which specifically addresses the title / question.

• decide what is relevant, necessary and appropriate for the task and their focus.

• select the main ideas and subordinate ideas that are to be presented to support their focus.

• order the ideas to present a logical viewpoint that is justified and presents a specific answer / conclusion.

- revise their plans / outlines following consideration and feedback (use sources)

- select suitable information from texts for inclusion in their writing

- distinguish between information to be summarised, paraphrased or quoted

- summarise, paraphrase or quote with appropriate referencing

- contextualise a quotation / re-presentation in their writing

- use a consistent reference format (writing)

- clearly indicate a specific focus in their writing and maintain that focus throughout the   essay

- present a clear argument / position throughout their writing – move beyond mere reporting / description / narration

- structure their essay to present their ideas clearly and in a logical sequence

- present a fully developed argument (utilise definitions, explanations, examples,   summaries, analysis etc. where appropriate)

- structure paragraphs such that they have a clear purpose, are fully developed and link logically to previous and subsequent paragraphs

- avoid factual inaccuracies / non-sensical and illogical statements / over-generalisations in   their writing

- avoid irrelevancies / unnecessary repetition of ideas in their writing

- avoid plagiarising other people's ideas / writing through an appropriate use of citation (post-writing)

• review / edit their own writing objectively

•   discuss   their   writing   with   others   and   consider   feedback

• revise their writing following feedback

Linguistic Accuracy and Document Formatting

By the end of ENG 101, students should be able to:

• demonstrate better fluency in the use of academic style and register

•   demonstrate   a   broader   range   of   general   academic   vocabulary

• show improvement in problematic areas of grammar:

- use of articles

-use of prepositions

- use of tenses

- subject-verb, gender, number agreement

- sentence fragments (sentences with no verb / subject – subordinate clauses presented as sentences eg. Because ……)

• show improvement in the mechanics of writing, in particular, spelling and punctuation

• present assigned work in a suitable fashion ie. A/4 paper - typed (or if handwritten, readable handwriting). - double-spaced - margins - clear, necessary information - biography

APPENDIX B

ENG 101 WRITING EXAM GRADING CRITERIA

	CONTENT	ORGANIZATION	LANGUAGE	CITATION/TECH
A 90-100	• answers the question set clearly and thoughtfully • presents a strong argument / progression of ideas that is convincing for the reader • has a clearly maintained focus that guides the reader well throughout	• has an engaging introduction that states the subject and the purpose of the essay • has well-structured, well-linked paragraphs that have one main idea and strong supporting ideas • has a powerfully convincing conclusion	• has very accurate language – only minor errors • impressive use of academic style and academic level vocabulary	• clear acknowledgment of writer and text throughout • clear emphasis on analysis and comment on the source rather than summarising • uses reference format with dexterity
B 78-89	• answers the question set clearly and in detail • presents an argument / a progression of ideas that is clear for the reader • has a clearly maintained focus	• has a well-written introduction that states the subject and the purpose of the essay • has well-structured paragraphs that have one main idea and good supporting ideas • has a clear conclusion which summarizes the main points and answers the question	• has reasonably accurate language – infrequent and fossilized errors • accurate use of academic style and good use of academic level vocabulary	• mostly refers to the writer and text accurately and appropriately • has a suitable balance of source material and analysis / comment • uses accurate references for source material that is quoted / used
C 70-77	• answers the question set • presents an argument / a progression of ideas that can be understood • has a reasonably maintained focus	• has an introduction that states the subject and the purpose of the essay • has separate paragraphs that have one main idea and supporting ideas • has a conclusion which summarizes the main points	• has language that is accurate enough not to impede meaning • shows some evidence of accurate use of academic style and vocabulary	• refers to the writer and text accurately • has a reasonable balance of source material and analysis / comment • source material that is used is referenced but sometimes inaccurately

.... PASS LEVEL ....

	CONTENT	ORGANIZATION	LANGUAGE	CITATION/TECH
C- 67-69	• answers the question set but underdeveloped and with some irrelevance • argument / progression of ideas is unclear • at times, may lose focus	• has an introduction in which the subject and the purpose of the essay are stated but not clear • has some paragraphs that have more than one main idea and weak supporting ideas • has a conclusion with little detail / unclear	• has language in which inaccuracy impedes meaning at times • shows little evidence of academic style and vocabulary	• may refer to the writer and text but inaccurately or inappropriately • more summary of source than analysis / comment • most source material is referenced but inaccurately
D 55-66	• fails to really answer question set • argument / progression of ideas is very difficult to follow • focus is not clear	• has an introduction which serves little purpose • has poor paragraphing – too long or short – main ideas and supporting ideas confused • has a conclusion which serves little purpose	• has language in which inaccuracy frequently impedes meaning – understanding takes effort • more general / basic English than academic level English	• does not refer to the writer or the text properly • mostly summary of source – little if any analysis / comment • inconsistent and poor referencing throughout
F 0-54	• does not answer the question set • no real argument or progression of ideas • no real focus	• no introduction • little attempt to organize into paragraphs – very confused • no conclusion	• language blocks meaning – evidence of sentences which cannot be understood • very basic, inaccurate English	• does not refer to the writer or the text at all • inaccurate summary of source • no source material is referenced – plagiarism is a problem

APPENDIX C

ENG 101 WRITING EXAM GRADING CRITERIA EVALUATION QUESTIONNAIRE

Dear Colleague,

This questionnaire has been designed as a research tool form ay MA thesis to enable met o collect data about the perceptions of FAE instructors of the common ENG 101 writing exam grading criteria. Your responses to this questionnaire will be invaluable in the evaluation of the grading criteria and in any subsequent adjustments or improvements. All responses will be treated confidentially. I would be really grateful if you could spare some time to fill in this questionnaire. I would like to thank you in advance for your co-operation and contribution.

Yeşim TARKAN
Office Phone : 290 51 56
E-mail: tyesim@bilkent.edu.tr

A. **Please circle the one appropriate for you.**

i) Gender   M   F

ii) I am a……… a) native speaker of English   b) non-native speaker of English

iii) I have a (n)………
   a) BA in _____   b) MA in _____   c) PhD in _____.

iv) How long have you been working in FAE (FYE/FAST/ELSU/FAE)? ---------- years.

iv) I am planning to conduct interviews as I would like to collect further more detailed feedback about the ENG 101 Writing Exam Grading Criteria. Of you would like to volunteer to take part in the interviews, could you pleas fill in the form that will be distributed along with the questionnaire?

THANK YOU VERY MUCH !

B. **In this section please refer to the attached ENG 101 Writing Exam Grading Criteria and indicate your personal opinions about it by ticking the relevant boxes below and adding any relevant comments you may have.**

SA: Strongly Agree   A: Agree   D: Disagree   SD: Strongly Disagree

		SA	A	D	SD	COMMENTS (if applicable)
1	The common ENG 101 Writing Exam Grading Criteria are good for making students' writing objectively across the FAE program.					
2	By having the common ENG 101 Writing Exam Grading Criteria, instructors across the FAE program have a common understanding of the ENG 101 course writing objectives.					
3	The common ENG 101 Writing objectives to be covered in the ENG 101 course.					
4	Using the common ENG 101 Writing Exam Grading Criteria saves time while marking.					
5	The common ENG 101 Writing Exam Grading Criteria ere user-friendly.					
6	The common ENG 101 Writing Exam Grading Criteria are good at discriminating between border-line fail papers.					

		SA	A	D	SD	COMMENTS (if applicable)
7	Having the common ENG 101 Writing Criteria enhances fair marking of students papers across the FAE program.					
8	Using the common ENG 101 Writing Exam Grading Criteria instructors can decide easily whether a student's paper passes o fails.					
9	In the criteria, the categories-content-organization-language and citation/tech-match the ENG 101 course writing objectives.					
10	It is good that all categories-content-organization-language and citation/tech in the criteria have equal weight while assessing students' written products.					
11	The descriptors (the bullet points for each category in each band) reflect the ENG 101 course writing objectives.					
12	The descriptors discriminate well between bands and within each band.					
13	The descriptors under each category in each band let me assign the appropriate grades easily to students' papers.					
14	The wording in the descriptors is easy to understand.					
15	I am confident about using the common ENG 101 Writing Exam Grading Criteria appropriately.					
16	I am confident that all of the raters/instructors across the program have a common understanding of the common ENG 101 Writing Exam Grading Criteria.					
17	I am satisfied with the common ENG 101 Writing Exam Grading Criteria.					
18	I am satisfied with the common ENG 101 Writing Exam Grading Criteria.					

C. **Please note down any further comments you may have about the strengths and weaknesses of the ENG 101 Writing Exam Grading Criteria.**

STRENGTHS ☺☺☺	POINTS TO BE CONSIDERED ☹☹☹

D. **Please note down any suggestions you might have to improve or revise the ENG 101 Writing Exam Grading Criteria.**

_____
_____
_____
_____
_____

THANK YOU FOR YOUR TIME AND COOPERATION….

APPENDIX D

INTERVIEW GUIDELINE

1. Why do you think the new ENG 101 writing exam grading criteria have been introduced?

2. What do you think about using a set/ standard writing exam grading criteria across the FAE program to assess students' written products?

3. What can you say about the descriptors in each band?

4. Is the wording in the descriptors clear or confusing? Please specify

5. In what ways do the bands match or not match the ENG 101 course objectives?

6. To what extent does ENG 101 writing exam grading criteria reflect actual teaching done in class?

7. What do you think about the equal weighting of the categories in the criteria?

8. Were there any sections of the criteria which you paid more or less attention to while marking?

9. Do you feel confident that there is a common understanding of the criteria among all the raters / instructors across the program? If yes, how do you know? If not, what could be done to overcome this problem?

10. Have you had any training on how to use the ENG 101 writing exam grading criteria? If not, would you like some? If yes, what kind of training?

11. What are the strengths of the ENG 101 writing exam grading criteria?

12. What are the points to reconsider for the ENG 101 writing exam grading criteria?

APPENDIX E

SAMPLE INTERVIEW TRANSCRIPTION

Date: 14.06.2005

1. Why do you think the new ENG 101 writing exam grading criteria

   have been introduced?

The why is pretty obvious, I think. I think we do need something in common to assess students writing. We need to have a common understanding of what is expected from the students and what we aim for in the course. Having such criteria is an advantage for both teachers and students as expectations are clear for both parties. Also, this leads to standardization across the program.

2. What do you think about using a set/ standard writing exam grading criteria across the FAE program to assess students' written products?

I agree with the idea of having set criteria across the program as I am new to the program and this helped me a lot. I think in the past each teacher had their own criteria and this caused some problems. To have such criteria was nice.

3.What can you say about the descriptors in each band?

Descriptors are done fairly well. However, few need to be tuned. One thing I like about these criteria is descriptors are continued through out all. For example, in 'A' band it says 'answers the question set clearly and thoughtfully' and in 'B' band it says 'answers the question set clearly and in detail.' So, this is good.

4. Is the wording in the descriptors clear or confusing? Please specify.

In general, they are clear. There are no perfect criteria. They are easy to understand.

5.In what ways do the bands match or not match ENG 101 course objectives?

This reflects most of the objectives. It cannot reflect all. This is impossible. For example, we focus on how to write an introduction and state the purpose of the essay. If you look at the criteria, under the heading of 'organization' in each band, the students' performances are evaluated.

6. To what extent ENG 101 writing exam grading criteria reflect actual teaching done in class?

It does come across. I mean, when I look at the final exam of the students, I do see what the students have achieved and their level. For example, I can see if the students got the idea of how ideas in an essay should flow or else.

7. What do you think about equal weighting of the categories in the criteria?

Personally, I do not agree that categories have equal importance. I believe, 'citation/ tech' should be lowered. When the exam task is considered, the students are not really referencing to the text. For me, spelling is difficult. I cannot spell. Students should not be penalized for that. For me, 'content' and 'organization' are more important. You can always fix the language and / or punctuation. Look at the teachers. How many of us can write very well in our first drafts? We always change things. I would be very upset if I was graded in this way.

8. Were there any sections of the criteria which you paid more or less attention to while marking?

As I mentioned earlier, I pay more attention to 'content' and 'organization'. I will use the same analogy that I use with the students here. Suppose that you have two different cars. A Porsche that does not run very well and the cheap Honda but it runs very well. You can always fix the citation or language but you cannot add anything to the content or organization easily. So you should write like a Porsche. Content and organization need to be better. Ok, maybe it does not run very well but you can fix it.

9. Do you feel confident that there is a common understanding of the criteria among all the raters / instructors across the program? If yes, how do you know? If not, what could be done to overcome this problem?

In our unit yes but not across the program. In our unit our grades are close but not with the others. We are crowded and teachers have different backgrounds. They teach differently. They grade differently. I believe this requires strong control organization.

10. Have you had any training on how to use ENG 101 writing exam grading criteria? If not, would you like some? If yes, what kind of training?

We had a sort of standardization but it was not really helpful. We did not walk out standardized. We need more standardization sessions.

11. What are the strengths of ENG 101 writing exam grading   criteria?

To sum up, I can say it is a sort of guide. It makes our expectations clear. It is easy to use. It helps for standardization across the program.

12. What are the points to reconsider for ENG 101 writing exam grading criteria?

Grades in the bans are confusing. I do not use grades. We should get rid of them. Content and organization should have more weighting. F band is a bit problematic as the range is from 0 to 54. Apart from those, I am happy with these criteria.