

Gender Differences in Item Format and Skill Area: Some Evidence from an EFL Achievement Test

Ayşe Engin*, Deniz Ortaçtepe**

Abstract

The present study investigated the extent to which male and female language learners' scores on achievement tests vary according to item format and skill areas. The statistical analysis of data from one achievement test administered to 303 pre-intermediate level students indicated that males and females' scores showed differences in both item format and skill areas. While females outperformed males significantly with 'find the correct form' and 'paragraph writing' questions, males did not show any superiority in any item format. Females also outperformed males in three skill areas; 'writing' 'grammar' and 'vocabulary' while males scored higher only in 'listening'. Possible reasons behind these differences between males and females' scores can benefit future researchers, language teachers, and administrators in terms of theoretical and practical perspectives.

Keywords: gender differences, item format, skill areas, achievement tests

© Association of Gazi Foreign Language Teaching. All rights reserved

1. Introduction

1.1 Gender differences in language learning

The achievement scores of language learners on language tests could be affected by students' proficiency levels as well as some other construct-relevant or construct irrelevant factors. One such variable is gender, which, as a broad term, is often used to denote not only the biologically based, dichotomous variable of *sex* (that is, male or female) but also the socially constructed roles (i.e. gender) created by "the ways sexes are raised and socialized within a certain culture" (Nyikos, 2008, p. 73). Although factors related to exams such as validity, reliability, practicality, (e.g., Fulcher & Davidson, 2007; Harris & McCann, 1994; Hughes, 2003) and features of test takers such as their attitude, motivation, and aptitude (e.g., Dörnyei, 2001; Genesee, 1976; Opler, 1989) have been widely discussed, gender as a variable has received little attention in the field of second language learning and teaching (Catalan, 2003; Nyikos, 2008; Sunderland, 1994). In fact, according to Graham (1997), of all the factors that influence test outcomes, gender is the one to which the least attention has been paid.

* Kadir Has University, Istanbul, Turkey. E-mail: ayse.engin@kadirhas.edu.tr

** Bilkent University, Ankara, Turkey. Email: deniz.ortactepe@bilkent.edu.tr

The biological hypotheses explaining the differences between males and females include three interrelated biological features: genetic, hormonal and brain (Halpern, 1992). First, genetic differences hypothesis proposes the theory that males and females have different intellectual abilities because they inherit different genetic codes, which may result in performance differences. For instance, according to Legato (2005), women have more nerve cells in the left part of the brain where language is centered. Second, the effect of hormones on the development of cognitive abilities has also been the focus of many studies (e.g., Gipps & Murphy, 1994; Halari et al., 2005). While one theory links “early physical maturation with intellectual development in order to explain girls’ assumed superiority in early language related skills” (Gipps & Murphy 1994, p. 58), another theory proposed that late maturers at puberty (typically males) exhibit “more highly developed spatial skills than verbal skills, whereas for early maturers (typically females) the converse is true” (Gipps & Murphy, 1994, p. 58). The third biological feature regarding gender differences is the brain. Halpern (1992) proposed that males and females differ in the way their brains are organized for intellectual behaviors. For instance, the female brain is less lateralized than the male brain, enabling language functions to be represented in both hemispheres (Levy, 1976).

There are also environmental theories regarding gender differences in linguistic performance. Wilder and Powell (1989) discussed the different ways males and females are encouraged to interact with the environment and the people around them. According to Gipps and Murphy (1994), there are expectations for girls to perform better in language domains than quantitative domains, and children’s judgments closely reflect those of their teachers and parents. Different approaches to different sub-groups (i.e. males and females) may encourage different skill development; in particular, boys are encouraged to develop independent, self-confident behaviors which are required more for future achievement in mathematics and science (Gipps & Murphy, 1994). On the other hand, Nyikos (2008) argued that adults have a subconscious perception of females’ language superiority, and talk more to baby girls than boys, respond more to girls’ early attempts to talk, and have longer, more complex conversations with daughters.

According to Wilder and Powell (1989), item content is also a source of differential performance because familiarity with content varies depending on the different life experiences of males and females. Students’ attitudes towards particular contents and subjects might also influence their performance. Males are more inclined to perceive mathematics and science as more valuable for their lives; however, both boys and girls think these content areas are not that important for girls (Gipps & Murphy, 1994). Therefore, such perceptions not only affect motivation but also influence their engagement with certain subject areas. Sunderland (2000) pointed out that a wide range of language-related phenomena (e.g., language learning styles and strategies) have been influenced by gender differences, thus, it is inevitable to expect a gender effect on language learning performance.

2. Review of Literature

There are several studies confirming the results that gender may cause differential performance, as well as other studies, which found no differences between males and females in foreign language skills. Yet, each of these studies focused on a different variable rather than comparing the overall language performance of males and females. For instance, Feyten (1991) did not find any differences in general language learning skills of male and female foreign language learners. Likewise, Bacon (1993) and Markham (1988) looked at the listening comprehension abilities of foreign language learners and could not identify any gender differences. Nyikos (1990), in a study on gender effect, found no difference in males’ and females’ rote memorization skills in terms of foreign language learning. As far as speaking is concerned, Farhady’s (1982) study revealed that females were better at recognizing the constituents of more or less prestigious dialects; thus, were able to differentiate among dialects better than males.

Bensoussan and Zeidner (1989) also found that in oral language tests, males not only reacted less negatively but also experienced less anxiety than their female counterparts.

A majority of research looking into gender differences examined females and males' performance in vocabulary. For instance, according to the results of Catalan's (2003) study, female language learners used a higher number of vocabulary learning strategies than males. She also found that 'analysis of the parts of speech for a new word' is reported as a more preferred strategy by females, while males preferred to analyze affixes and roots. Gu's (2002) study also revealed that female participants outperformed male participants on both vocabulary size and general English proficiency. In another study conducted by Boyle (1987), female Chinese learners were stronger in overall language ability while male learners of English in China were found to be stronger in terms of vocabulary recognition in a listening task.

On the other hand, there are also a few studies that examined the differences between males and females in terms of socio-affective factors. Politzer (1983), for instance, investigated the social behavior of language learners, and found that social strategies were used more by females. In Politzer's (1983) study, females expressed more interest in interpersonal relationships (e.g., cooperativeness) and less interest in competitiveness and aggression. Oxford and Nyikos (1989) also found that formal rule-based strategies, general study strategies, and conversational input-elicitation strategies were used more often by females than males.

Ryan and Demark (2002) addressed the issue of gender differences through two related meta-analytic studies of published and present research. The analysis of students' achievement in language assessments suggested that females outperformed males in language assessment if a constructed-response format (e.g., short answer, essay) was employed, but not when their language skills were measured with selected-response items (e.g., multiple choice, true/false, matching). This result reflects a gender difference favoring females in writing performance scores. It also implies that, as a result of item format, there might be differences of achievement between males and females in the skill areas as well. Females' success in constructed-response format questions suggests a higher achievement in writing skill compared to males. In Graham's (1997) study with German learners, students were asked about their opinions regarding different aspects of language. According to the results, male students felt less comfortable with reading than their female counterparts, but they felt more comfortable with oral work and general grammar. These differences in attitude may also result in differences of achievement according to the skill area tested; thus, affecting students' performance in language achievement tests.

In summary, while some studies showed an advantage for women in language learning (e.g., Gu, 2002; Sunderland, 2000), others report no significant relationship between gender and language learning (e.g., Ehrman & Oxford, 1995). Hence, the small number of studies, some of which date really back, indicate inconsistencies in the literature about the role or effect of gender on language learning. In order to obtain reliable test scores, the abilities test developers want to measure should be differentiated from the other factors that might affect the test-takers' scores such as gender, item format and skill areas. The present study, by comparing females and males' scores in an achievement test in regards to item format and skill areas, aims to address the following question; To what extent do male and female language learners' scores vary according to item format and skill areas?

Understanding the strategies employed by male and female students will help instructors guide their students. The results of this study will reveal the strategies used by males and females in order to answer the questions. By identifying the strategies and informing the students and training them to self-control, teachers can help students monitor their comprehension and learning processes better. Also, being aware of the variety of the strategies used by the students, teachers might be more sensitive to different learning styles and intelligences. If there are any strategies that work better for a particular skill, students might be trained to employ those strategies, or at least taught and given an option to select those strategies

depending on which skill they are required to use. By this way, teachers can take the steps to maximize the chances of success.

Variety of question types in any skill area is also important not to favor one group who is good at a particular question format. The topics of the stimulus texts in the exam should be also neutral; not appealing to either males or females in an obvious way. If females are better at memorizing, the questions requiring memorization should be minimized not to give an unfair advantage to the females. This study will reveal if either group is better at a particular question format, which will help the test developers to write unbiased tests not to favor males or females.

3. Method

3.1. Sample

This research was conducted at a private university located in Istanbul, Turkey. The sample for this study comprised pre-intermediate level students in the second module of the preparatory school which provides 26 hours of English instruction a week. It was purposeful sampling; pre-intermediate was selected due to the level with the highest number of enrolled students, and it was assumed that the higher number of participants would provide more reliable results. The sample included 303 students (163 males; 140 females). These students were young adults from different majors with different educational, social, and economic backgrounds.

3.2. Data source

The second achievement test of the second module administered to pre-intermediate level students was analyzed since it was the most recent test with more variety of questions types. The test was written by one of the members of the testing office with the guidance of an expert consultant holding two PhD degrees in the field of language testing. The test writer stated that the students were familiar with the question types and content of the test. The test included five sections; reading, writing, listening, grammar, and vocabulary with six different item formats (see Table 1).

Table 1. Skill areas and distribution of item formats

Skill Area	Selected-Response Questions	Constructed-Response Questions
Reading	matching find the definition	open-ended
Writing	none	write a paragraph
Listening	multiple choice fill in the blanks	none
Grammar	fill in the blanks find the correct form	none
Vocabulary	matching	none

All parts of the tests coming from 303 students and grading sheets belonging to all pre-intermediate classes were taken. All the tests were graded by two teachers randomly assigned for the task and discrepancies (if any) were checked after both graders marked the tests. The teachers assigned the final grades together. The grading scales and answer keys were written and provided by the test writer. The

researchers entered these grades into SPSS 17 by identifying the items in terms of their format; (matching, fill in the blanks, find the correct form, multiple choice, open ended, and paragraph writing) and the skill areas tested (reading, writing, listening, grammar and vocabulary). The gender of the students was coded according to the demographic information provided by the university.

4. Results

4.1. Gender differences in terms of item format

A two-way ANOVA test was conducted to identify whether gender has an effect on the achievement scores in different item formats (see Table 2). As presented in Table 2, the results showed that gender, as a variable, does not have a main effect on students' scores in different item formats; that is, being male or female does not influence students' performance in achievement tests.

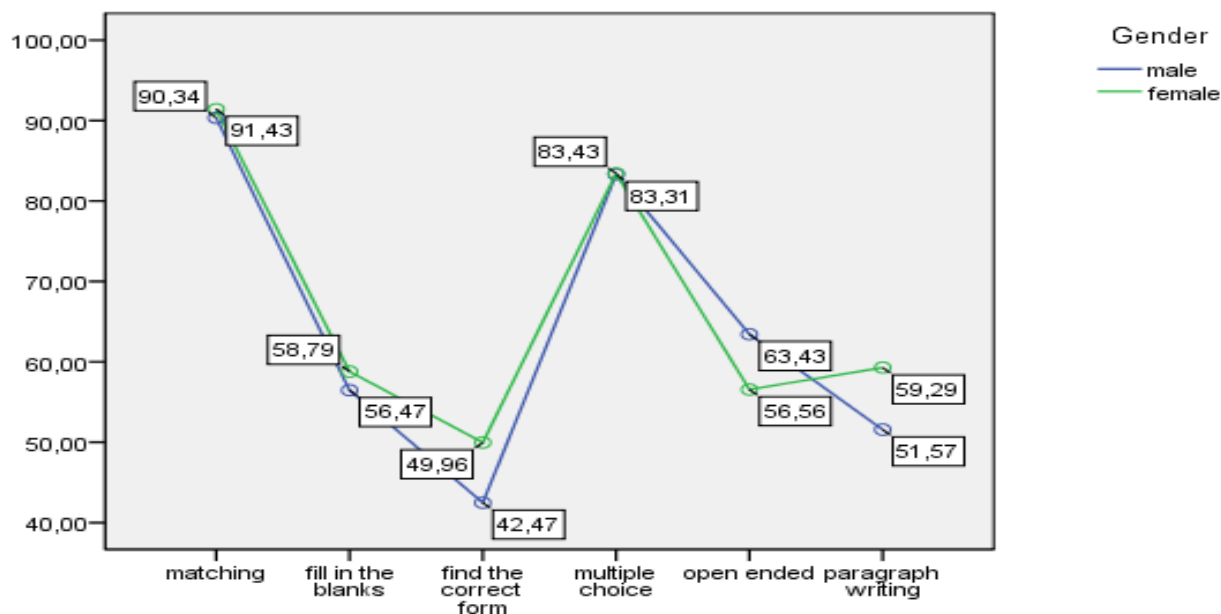
On the other hand, the main effect of item format on students' scores was statistically significant such that scores in certain item formats were significantly higher than the others ($F(5, 1806) = 166.392, p = .000, \eta^2 = .315$). More importantly, there was a highly significant interaction effect between gender and item format ($F(5, 1806) = 3.972, p = .001$). However, the eta squared statistic ($\eta^2 = .011$) indicated a small effect size which may be related to the lack of main effect of gender on the scores. Because of the small effect size of interaction between gender and item format, item format ($\eta^2 = .315$) alone explains more of the variation than the interaction effect (see Figure 1).

Table 2. Variation of students' scores according to gender and item format

Dependent Variable: Score						
Source	SS	df	MS	F	Sig.	Partial Eta Squared
Gender	1765.906	1	1765.906	3.189	.074	.002
Item Format	460666.757	5	92133.351	166.392	.000*	.315
Gender * Item Format	10997.074	5	2199.415	3.972	.001*	.011
Error	1000006.631	1806	553.714			
Total	9282594.000	1818				

a. R Squared = .324 (Adjusted R Squared = .320)

As shown in Figure 1, the interaction effect of gender and item format lies in 'open ended' questions where, unlike any other item format, males outperform females. Table 3 below details the difference between males' and females' scores according to item format.

Figure 1. Gender and item formats

As presented in Table 3 below, the biggest difference between the means of males' and females' scores was in 'paragraph questions' (6). The mean of males' scores for 'paragraph questions' was 51.57, while it was 59.28 for females, a statistically significant difference at the $p < .01$ level. The second highest difference between the means of males and females scores was in 'find the correct form' questions (3). The difference was 7.49; a statistically significant result at the $p < .05$ level, with a mean of 49.96 for females and 42.47 for males. 'Find the correct form' and 'paragraph writing' questions were the only item formats with a statistical difference between males' and females' scores.

The third highest difference between the means of males and females scores was in 'open ended' questions (5). This item format showed a different pattern because, as mentioned earlier, it is the only item format where males scored higher than females with \bar{x} male = 63.42 and \bar{x} female = 56.55. The fourth biggest difference between the means of males and females scores was in 'fill in the blanks' questions (2) with \bar{x} males = 56.46 and \bar{x} females = 58.78. There is a small difference in 'matching' questions (1), where both genders scored best (\bar{x} males = 90.33, \bar{x} females = 91.42). The smallest difference between the means of males and females was in the 'multiple choice' questions with \bar{x} female = 83.42, and \bar{x} male = 83.31. None of these differences, however, were statistically significant.

Overall, these results indicated that there was no main effect of gender on students' scores according to different item formats; however, item format did have a main effect on scores influencing students' success in different item formats. There is also an interaction effect between gender and different item formats, when these two variables interact, they do affect the scores students receive from different item formats.

Table 3. Comparison of mean scores of males and females in item formats

		Scores		T-test	df	Sig. (2-tailed)
		\bar{x}	SD	t		
1. Matching	Males	90.34	21.75	-.435	301	.664
	Females	91.43	21.78			
2. Fill in the blanks	Males	56.46	19.65	-1.036	301	.301
	Females	58.78	19.14			
3. Find the correct form	Males	42.47	27.40	-2.393	301	.017*
	Females	49.96	26.89			
4. Multiple Choice	Males	83.31	12.62	-.077	301	.939
	Females	83.42	13.55			
5. Open-ended	Males	63.42	56.55	1.767	301	.078
	Females		34.12			
6. Paragraph	Males	51.57	19.90	-3.257	301	.001*
	Females	59.28	21.29			

Note. * $p < .01$

4.2. Gender differences in terms of skill areas

A two-way ANOVA test was conducted to identify whether gender has an effect on achievement scores in different skill areas (see Table 4). As shown in Table 4, there was a significant main effect of gender on the scores received in different skill areas ($F(1, 1505) = 8.284, p = .004, \eta^2 = .005$). That is, being male or female influenced students' performance in different skill areas.

The main effect of skill areas on students' scores was also statistically significant such that scores in certain skill areas were significantly higher than the others ($F(4, 1505) = 14.926, p = .000, \eta^2 = .038$). More importantly, there was a significant interaction effect between gender and skill areas ($F(4, 1505) = 4.699, p = .001$). However, the eta squared statistic ($\eta^2 = .012$) indicated a small effect size. Although gender was significant, its effect size was really small, thus, the effect size for the interaction was also small. Figure 2 below shows the interaction effect between gender and skill areas.

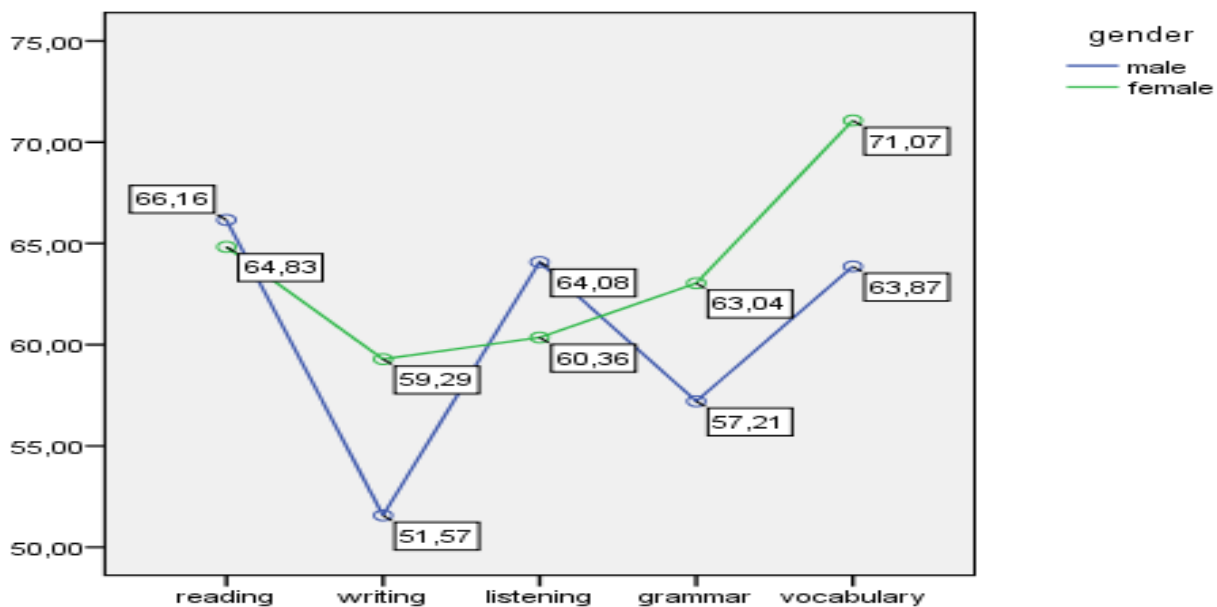
Table 4. Comparison of gender and skill areas

Dependent Variable: score						
Source	SS	df	M	F	Sig.	Partial Eta Squared
Gender	3710.506	1	3710.506	8.284	.004*	.005
Skill	26741.258	4	6685.315	14.926	.000*	.038
Gender * Skill	8418.024	4	2104.506	4.699	.001*	.012
Error	674096.900	1505	447.905			
Total	6542605.500	1515				

Note. * $p < .01$, a. R Squared = .056 (Adjusted R Squared = .050)

As shown in Figure 2, the interaction effect of gender and skill areas lies in 'listening' questions where unlike any other item format males outperform females significantly. Table 5 below details the difference between males' and females' scores according to skill areas.

Figure 2. Gender and skill areas



As shown in Table 5 below, the biggest difference between the means of males and females was in 'writing.' Females performed better than males in writing (\bar{x} female = 59.28 and \bar{x} male = 51.57), a statistically significant difference at the $p < .01$ level. The second highest difference between the means of males and females was in 'vocabulary.' There was again a female superiority with \bar{x} female = 71.07 and \bar{x} male = 63.86, another statistically significant difference at the $p < .05$ level.

The third highest difference between the means of males and females was in 'grammar.' Males' mean score in grammar was 57.20, whereas females' mean score was 63.03, a 5.83 difference that is statistically significant at the $p < .05$ level. The fourth biggest difference between the means of males and females was in 'listening,' in which males ($\bar{x} = 64.07$) performed better than females ($\bar{x} = 60.35$), and the difference in this skill area was statistically significant at the $p < .05$ level. Reading was the second skill where males performed a little better ($\bar{x} = 66.16$) than females ($\bar{x} = 64.83$), which was the smallest difference between the means of males' and females' scores, and also the only difference which was not statistically significant.

Overall, these results indicated that there was a main effect of gender on students' scores according to different skill areas. Also skills areas did have a main effect on scores influencing students' success in different skill areas. There was also an interaction effect between gender and different skill areas, when these two variables interact, they do affect the scores students receive from different skill areas.

Table 5. Comparison of mean scores of males and females in skill areas

		Scores		T-test		
		\bar{x}	SD	t	df	Sig. (2-tailed)
1. Reading	Males	66.16	22.29	.523	301	.601
	Females	64.83	21.84			
2. Writing	Males	51.57	19.90	-3.257	301	.001*
	Females	59.28	21.29			
3. Listening	Males	64.07	16.77	1.989	301	.048*
	Females	60.35	15.60			
4. Grammar	Males	57.20	20.20	-2.539	301	.012*
	Females	63.03	19.57			
5. Vocabulary	Males	63.86	26.83	-2.419	301	.016*
	Females	71.07	24.06			

Note. * $p < .01$

5. Discussion

5.1 General differences between male and female language learners

The female superiority in most skill areas and in the two significantly different item formats in the test confirmed the results of many other studies conducted in the field (e.g., Boyle, 1987; Nyikos, 1990). The significant differences between females and males can be attributed to several factors such as differences in cognitive abilities that result in different language learning strategies, biological factors and affective reasons.

Yang (2001) noted that there are gender differences regarding the development of cognitive abilities between males and females, which may result in different cognitive strategy preferences by males and females. These different cognitive strategies may favor one group in language learning, which turns out to be females in this study. Ehrman and Oxford (1995) listed those language learning strategies more often used by females as; metacognitive (e.g., higher order executive skills; planning, evaluating, organizing), affective (e.g., emotional and motivational) and social strategies (e.g. asking questions to get verification, asking for clarification of a confusing point). For instance, females are better at general study strategies, refer to rule-related strategies, and use conscious learning strategies more than males (Oxford & Nyikos, 1989). Thus, females in the present study may have surpassed their male counterparts because they employed more advanced and effective language learning strategies.

Biological factors might also have provided females with an advantage over males. According to Springer and Deutsch (1989), females were less lateralized for language functions, and were superior to men in language skills. Some researchers have also related performance differences between males and females regarding aptitude. Powell (1979) claimed that females are superior to males in all aspects of

linguistic process; therefore, show a greater aptitude for language. Rua (2006) also stated that “although both males and females have the same linguistic potential as human beings (aptitude in general sense), females’ linguistic skills somehow seem more prone to be stimulated in order to reach higher levels of linguistic competence” (p. 103).

Another reason why females outperformed males in three major skill areas in the present study’s achievement test may be related to affective reasons. Ellis (1994) recognized attitude as the obvious explanation for females’ greater success in second language learning. On the other hand, Spolsky (1990) looked at the relationship between attitude and success in language learning from a different perspective. He suggested that “attitudes do not have direct influence on learning, but they lead to motivation, which does” (p. 49). The studies conducted by Pritchard (1987), Powell and Littlewood (1983), and Powell and Batters (1985) showed that unlike males, females were more favorably inclined to the language itself, the speakers, and the cultures of other languages. Krashen (1988) regarded integrative motivation as a stronger predictor of achievement than instrumental motivation. Socialization also seems to determine, or at least, influence motivation, and also cognitive development (Slavin, 1988). Social forces such as parental attitude and gender related cultural beliefs determine how males and females perceive the process of language learning, and the value they attach to other languages (Oxford, 1993; Schmenk, 2004). Hence, in a culture where language is seen as a women’s topic, it is natural to find out female superiority in language learning.

5.2 Gender, Item Format and Skill Areas

In the present study, the males and females’ scores in different item formats as well as in different skill areas in the achievement test were analyzed. The results revealed that males’ and females’ scores varied significantly in the four skill areas; namely ‘writing,’ ‘listening,’ ‘grammar’ and ‘vocabulary.’ In ‘writing,’ ‘grammar’ and ‘vocabulary,’ females outperformed males; however, only in ‘listening’ males scored higher than females. Additionally, males’ and females’ scores varied significantly in two item formats, ‘find the correct form’ and ‘paragraph writing’ questions. In both item formats, females scored significantly higher than males. Yet, most studies focusing on gender differences investigated the differences in skill areas and strategies used (e.g., Bacon, 1993; Boyle, 1987; Catalan, 2003; Gu, 2002) rather than item formats. In this respect, the findings of this study, regarding the difference in the item formats, might also be explained in reference to the skill areas where these item formats were used.

Paragraph writing questions and the writing skill. As far as the ‘paragraph writing’ question, which cannot be evaluated separately from writing skill, is concerned, the fact that females outperformed males was an expected result (e.g., Graham, 1997; Knudson, 1995). Females not only are better writers (Knudson, 1995) but also more frequently use a careful, organized approach (i.e., planning, monitoring and evaluating) than males, who expressed a dislike for planning as a strategy in writing (Graham, 1997). This more planned and organized approach utilized by females in writing might have favored the female language learners in achieving higher scores in paragraph writing questions.

On the other hand, there are also studies with contradictory findings. Pajares and Valiente (1996) found no differences between males’ and females’ writing performances in their study; however, they reported that females had higher writing self-efficacy, that is, “beliefs about their capabilities to produce designated levels of performance” (Bandura, 1994, p. 2). Shell (1989) investigated the writing self-efficacy of undergraduate students and found a strong relationship between students’ confidence in their writing skills and their grades in holistic scales. There are numerous studies which report that females express stronger self-beliefs in language arts than males (e.g., Eccles, Wigfield, Flanagan, Miller, Reuman, & Yee, 1989; Pajares, Miller, & Johnson, 1999; Pajares & Valiante, 1997). Thus, the differences between males’ and females’ achievement in the present study might be related to students’ self-efficacy or confidence in

writing in the sense that the females might have had higher self-efficacy (Eccles et. al., 1989) or more confidence (Pajares & Valiante, 1997; 2001) than the males which might have had a positive effect on their writing scores.

Females having stronger confidence in their writing capabilities than males (Pajares & Valiante, 1997) might also affect their ability to employ various self-regulatory strategies such as self-observation, self-evaluation, and self-correction while writing. Employing these strategies might result in higher achievement in the writing section of the test. Another reason that might have led the present study's female learners to perform better in writing could be related to writing apprehension. Daly and Miller (1975) who first used writing apprehension to describe a form of writing anxiety, reported strong correlation between apprehension and perceived likelihood of success in writing. They also found that males were more apprehensive about writing than females. Similarly, the male students in the present study might have experienced more apprehension towards writing and this might have influenced their achievement in writing negatively.

Find the correct form questions and grammar. 'Grammar' is the second skill area in which females scored significantly higher than males. Thus, females also outperformed males in 'find the correct form' questions in which grammar was tested. In a review article, Oxford, Nyikos and Ehrman (1988) argued that women remembered more details, which might explain why females scored higher in 'find the correct form' questions. Remembering the correct verb form is essential to answer the 'find the correct form' questions in the test because all the questions in this format were about verb tenses. If females are better at remembering things, then this ability gives them a great advantage over males in question types such as 'find the correct form' because the memory factor is really influential. Also, females are more likely to learn grammatical items "by heart" (Graham, 1997, p. 81), which confirms Oxford, Nyikos and Ehrman's (1988) assumption that females are better at remembering things.

On another note, Oxford and Nyikos (1989) argued that women had a greater tendency for social approval and this tendency motivated them to strive for higher grades than men (Oxford, Nyikos & Ehrman, 1988). Confirming this argument, Van Houtte (2004) claimed that females' culture is more study-oriented and supportive of academic achievement. If so, females might have studied harder than males to memorize the verb forms, and learned the previously taught grammar topics better, and thus, received higher grades in 'find the correct form' questions.

Graham (1997) found out that male learners of German were more comfortable about grammar than females. In the same study, female learners of French expressed greater worries for grammar. Hence, even though Graham (1997) had participants who were learning two different languages, the findings are consistent in the sense that the male participants felt more comfortable with grammar. However, in the present study, the fact that the males underperformed in comparison with the females in 'grammar' where 'find the correct form' questions are asked, might imply that the difference did not result from the skill area tested but the item format, find-the-correct form, which requires rule-practice. According to Oxford and Nyikos (1989), females more frequently use formal rule-related practice than males. Hence, rule-related practice obviously provides an advantage in grammar which prescribes specific rules to be followed. The females in this study, as indicated in the literature, may have executed more appropriate strategies such as rule practice while studying for grammar, which might have favored them in grammar.

Vocabulary. Vocabulary is the last skill females scored higher than males. The findings regarding higher achievement of females in 'vocabulary' was confirmed by Gu's (2002) study in which females reported more use of almost all the strategies (e.g., guessing, using contextual clues, taking notes, and employing oral repetitions) that are associated with success in English language learning (Gu & Johnston, 1996). Catalan (2003) also found out that females used a higher number of vocabulary strategies than males. In Gu's (2002) study, females were found to have spent significantly more extra-curricular time on learning English than their male counterparts. When combined, employing more effective strategies and

investing more in language learning by spending more time are the factors that might explain female superiority in 'vocabulary.' In Oxford, Lavine, Hollaway, Felkins and Saleh's (1996) study, females have been found to try out new techniques for vocabulary learning. In that sense, different preferences in the number and type of vocabulary learning strategies might have caused a difference in the students' achievement in vocabulary.

Listening. In the present study, the results revealed a different pattern in the 'listening' skill where the males significantly outperformed the females. This finding conflicted with the studies of Markham (1988) and Bacon (1993) who found no difference in the listening comprehension of males and females. The findings of this study regarding males' superiority in 'listening' also contradicted the findings of the study conducted by Farhady (1982) who noted that females had higher comprehension in listening than males. Eisenstein (1982) argued that females discriminated dialects and prestige of dialects better than males. Different from these studies, Boyle (1987) revealed that males were better in recognizing words in 'listening' texts, which may bring higher comprehension; however, he also reported that females were better in general 'listening' comprehension (as cited in Kunnan, 1998). Larsen-Freeman and Long (1991) argued that females were better in listening. The findings of the present study seem to contradict the aforementioned studies in the sense that, instead of females, males performed better in 'listening.'

The reason why the present study contradicts the other studies might lie in affective factors such as anxiety, confidence, or motivation. Bacon's (1992) study on listening strategies indicated that males were more confident in their ability to tackle an aural passage might support this assumption. The difference might be also related to the listening strategies employed by males and females. In her study, Bacon (1992) reported that females used both metacognitive and cognitive strategies; strategies that manipulate information such as summarizing or reorganizing. Females adjusted metacognitive strategies according to the difficulty of the passage, but they used cognitive strategies in the same fashion without adjusting, even when they listened to more difficult passages. Males in the same study dealt with difficult texts more aggressively with reference to bottom-up strategies. In Graham's (1997) study, males were more likely to use problem-identification, a more direct, even confrontational strategy. Hence, strategy preferences of males and females in the achievement test might have resulted in different achievement scores favoring males. If the superiority of males stems from their use of strategies, it means that the strategies they employ seem to be effective.

6. Conclusion

This study yielded both expected and unexpected results in terms of the differences between males' and females' scores in skill areas and item format in an achievement test. These results might have been caused by a number of different reasons or interaction of these reasons. While some of the findings confirm the literature on gender studies in language testing, others contradict. The findings might be supported or contradicted by further research comparing genders in their achievement in terms of item formats and skill areas. Like all studies, the current study has a number of limitations, each of which provides an opportunity for future research. First, future studies can investigate the gender differences at different proficiency levels because students at different proficiency levels might tend to succeed differently in different item formats or skills. Second, because the achievement test analyzed in this study lacked a speaking part, gender differences in the speaking skill can be the focus of further studies.

Lastly, based on the findings of this research, it is recommended that language teachers and test-developers train language learners for various strategies of test-taking so as to minimize the differences in achievement of males and females and include a variety of question types in any skill area, as well as topics and themes in order not to favor one group who is good at a particular question format or familiar with a certain topic.

Biostatement for the first author:

Ayşe Engin graduated from Boğaziçi University, Foreign Language Teaching department in 2006. She earned Cambridge ESOL DELTA (Diploma in Teaching English to Adults) in 2011. She completed her M.A. degree on TEFL at Graduate School of Education, Bilkent University in 2012. She is working at Kadir Has University Preparatory School as Curriculum and Testing Unit Officer. Her research interests are second language acquisition, language testing, curriculum design and gender differences in second language learning.

Biostatement for the second author:

Dr. Deniz Ortaçtepe completed her BA (in ELT) and MA degrees (in Educational Sciences) at Boğaziçi University, Turkey. After working in Turkey for several years both as a research assistant and as an English teacher, in 2007 she moved to the United States to pursue her doctoral degree. In May 2011, she received her doctorate degree in Curriculum and Instruction at State University of New York-Albany, where she was also teaching academic writing to graduate students. Her research interests are second language socialization, teacher identity, intercultural pragmatics, and sociolinguistics.

References

- Bacon, S. M. (1992). The relationship between gender, comprehension, processing strategies and cognitive and response in foreign language listening. *The Modern Language Journal*, 76(2), 160-176
- Bacon, S. M. (1993). The relationship between gender, comprehension, processing strategies and cognitive and affective response in foreign language listening. *Modern Language Journal*, 76, 160-178.
- Bandura, A. (1994). Self-efficacy. In V. S. Ramachandran (Ed.), *Encyclopedia of human behavior*, vol. 4, pp. 71-81). New York: Academic Press. (Reprinted in H. Friedman [Ed.], *Encyclopedia of mental health*. San Diego: Academic Press, 1998).
- Bensoussan, M. & Zeidner, M. (1989). Anxiety and achievement in a multicultural situation: the oral testing of advanced English reading comprehension. *Assessment and Evaluation in Higher Education*, 14(1), 40-54.
- Boyle, J. P. (1987). Sex differences in listening vocabulary. *Language Learning*, 37, 3273-3284.
- Catalan, R. M. J. (2003). Sex differences in L2 vocabulary learning strategies. *International Journal of Applied Linguistics*. 13(1), 54-77.
- Daly, J. A. & Miller, M. D. (1975). Apprehension of writing as a predictor of message intensity. *Journal of Psychology*, 89, 175-177.
- Dörnyei, Z. (2001). *Motivational strategies in the language classroom*. Cambridge: Cambridge University Press.
- Ehrman, M. E. & Oxford, L. R. (1995). Cognition plus: Correlates of language learning success. *Modern Language Journal*, 79(1), 67-89.
- Eisenstein, M. (1982). A study of social variation in adult second language acquisition. *Language Learning*, 32, 367-391.
- Eccles, J. S., Wigfield, A., Flanagan, C., Miller, C., Reuman, D., & Yee, D. (1989). Self-concepts, domain values, and self-esteem. Relations and changes in early adolescence. *Journal of Personality*, 57, 283-310.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Farhady, H. (1982). Measures of language proficiency from the learner's perspective. *TESOL Quarterly*, 16, 43-59.
- Feyten, C. M. (1991). The power of listening ability: An overlooked dimension in language acquisition. *Modern Language Journal*, 75, 173-180.
- Fulcher, G. & Davidson, F. (2007). *Language testing and assessment*. New York: Routledge.
- Genesee, F. (1976). The role of intelligence in second language learning. *Language Learning*, 26(2), 267-280.
- Gipps, C. & Murphy, P. (1994). *A fair test? Assessment, achievement and equity*. Buckingham: Open University Press.
- Graham, S. (1997). *Effective language learning: positive strategies for advanced level language learning*. Clevedon: Multilingual Matters.

- Gu, Y. & Johnson, R. K. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning*, 46, 643-679.
- Gu, Y. (2002). Gender, academic major, and vocabulary learning. *A Journal of Language Teaching and Research*, 33(1), 35-54.
- Halari, R., Hines, M., Kumari, V., Mehrotra, R., Wheeler, M., Ng, V., Sharma, T. (2005). Sex Differences and Individual Differences in Cognitive Performance and Their Relationship to Endogenous Gonadal Hormones and Gonadotropins. *Behavioral Neuroscience*, 119(1), 104-117.
- Halpern, D. F. (1992). *Sex differences in cognitive abilities*. Hillsdale, NJ: Lawrence Erlbaum.
- Harris, M. & McCann, P. (1994). *Assessment*. Oxford: Heinemann.
- Hughes, A. (2003). *Testing for language teachers*. UK: Cambridge University Press.
- Knudson, R. E. (1995). Writing experiences, attitudes, and achievement of first to sixth graders. *Journal of Educational Research*, 89, 90-97.
- Krashen, S. (1988). *Second language acquisition and second language learning*. Englewood Cliffs, NJ: Prentice Hall.
- Kunnan, A. J. (1998). Approaches to validation in language assessment. In A. J. Kunnan (Ed.), *Validation in language assessment* (pp. 1-16). Mahwah, N.J.: LEA.
- Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research*. New York: Longman.
- Legato, M. J. (2005). *Why men never remember and women never forget*. New York: Rodale.
- Levy, J. (1976). Cerebral lateralization and spatial ability. *Behavior Genetics*, 6, 171-188.
- Markham, P. L. (1988). Gender differences and the perceived expertness of the speaker as factors in ESL listening recall. *TESOL Quarterly*, 22, 297-406.
- Nyikos, M. (1990). Gender-related differences in adult language learning: Socialization and memory factors. *Modern Language Journal*, 74(3), 273-287.
- Nyikos, M. (2008). Gender in language learning. In C. Griffiths (Ed.), *Lessons from good language learners: Insights for teachers and learners* (pp.73-82). Cambridge University Press.
- Obler, L. K. (1989). Exceptional second language learners. In: S. Gass, C. Madden, D. Preston & L. Selinker (Eds.) *Variation in second language acquisition: Vol. II. Psycholinguistic Issues* (pp. 141-159). Clevedon: Multilingual Matters.
- Oxford, R., Nyikos, M. & Ehrman, M. (1988) Vive la difference? Reflections on sex differences in use of language learning strategies. *Foreign Language Annals*, 21, 321- 329.
- Oxford, R., & Nyikos, M. (1989). Variables affecting choice of language learning strategies by university students *Modern Language Journal*, 73(3), 291-300.
- Oxford, R. (1993). Research on second language learning strategies. *Annual Review of Applied Linguistics*, 13, 175-187.
- Oxford, R. L., Lavine, R. Z., Holloway, M. E., Felkins, G., & Saleh, A. (1996). Telling their stories: Language learner use diaries and recollective studies. In R. L. Oxford (Ed.), *Language learning strategies around the world: Cross cultural perspectives* (pp. 19-34). Manoa: University of Hawaii Press.
- Pajares, F. & Valiente, G. (1996). *Predictive utility and causal influence of the writing self-efficacy beliefs of elementary students*. Paper presented at the annual meeting of the American Educational Research Association. New York.
- Pajares, F. & Valiante, G. (1997). The predictive and mediational role of the writing self-efficacy beliefs of upper elementary students. *Journal of Educational Research*, 90, 353–360.
- Pajares, F., Miller, M. D., & Johnson, M. J. (1999). Gender differences in writing self-beliefs of elementary school students. *Educational Psychology*, 91, 50–61.
- Pajares, F. & Valiante, G. (2001). Gender differences in writing motivation and achievement of middle school students: A function of gender orientation? *Contemporary Educational Psychology*, 26, 366-381.
- Politzer, R.L. (1983). An exploratory study of self-reported language learning behaviors and their relation to achievement. *Studies in Second Language Acquisition*, 6, 54-68.
- Powell, R.C. (1979). Sex differences in language learning: a review of the evidence. *Audiovisual Language Journal*, 17(1), 19-24.
- Powell, R.C. & Littlewood, P. (1983). Why choose French? Boys' and girls' attitudes at the option stage. *The British Journal of Foreign Language Teaching*, 21(1), 36-39.
- Powell, R.C. & Batters, J.D. (1985). Pupils perceptions of foreign language learning at 12+: some gender differences. *Educational Studies*, 11(1), 12-23.
- Pritchard, R. M. O. (1987). Boys' and girls' attitudes towards French and German. *Educational Research*, 29(1), 65–72.
- Rua, P. L. (2006). The sex variable in foreign language learning: an integrative approach. *Porta Linguarum*, 6, 99-114.

- Ryan, J. M., & DeMark, S. (2002). Variation in Achievement Scores Related to Gender, Item Format, and Content Are Tested. In G. Tindal & T. Haladyna (Eds.), *Large Scale Assessment Programs for All Students: Validity, Technical Adequacy, and Implementation* (pp. 67-88). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Schmenk, B. (2004). Language learning: a feminine domain? The role of stereotyping in constructing gendered learner identities. *TESOL Quarterly*, 38(3), 514-538.
- Shell, D.F. (1989). Self-efficacy and outcome expectancy mechanisms in reading and writing achievement. *Journal of Educational Psychology*, 81, 91-100.
- Slavin, R.E. (1988). Cooperative learning and student achievement. In R.E. Slavin (Ed.) *School and Classroom Organization*. Hillsdale, NJ: Erlbaum.
- Spolsky, B. (1990). *Conditions for second language learning*. Oxford: Oxford University Press.
- Springer, S. & Deutsch, G. (1989). *Left brain, right brain*. New York: Freeman.
- Sunderland, J. (1994). *Exploring gender: Questions and implications for English language education*. Hemel Hempstead: Prentice Hall.
- Sunderland, J. (2000). Issues of language and gender in second and foreign language education. *Language Teaching*, 33, 203-223.
- Van Houtte, M. (2004). Why boys achieve less at school than girls: the difference between boys' and girls' academic culture. *Educational Studies*, 30(2), 159-173.
- Wilder, G. Z. & Powell, K. (1989). *Sex differences in test performance: A survey of the literature*. New York: College Board Publications.
- Yang, Y. L. (2001) Sex and language proficiency level in color-naming performance: an ESL/EFL perspective. *International Journal of Applied Linguistics*, 11(2), 238-56.