UNIVERSAL WISER
PUBLISHER

Research Article

Easier and More Enjoyable: Recognizing Mistakes by English Foreign Language (EFL) Learners Across Gender

Neda Moezzipour [iD], **Jassem Fathabadi**[*] [iD]

Department of English Language and Literature, Ferdowsi University of Mashhad, Iran
Email: jassem_fathabadi@yahoo.com

Abstract: Daniel Kahneman asserted that recognizing others' mistakes is both easier and more enjoyable. The current study aimed at testing this claim in the context of English language learning among Iranians with regards to grammar tests. To this end, 150 Iranian advanced English as a Foreign Language (EFL) learners (75 males and 75 females) took a grammar test consisting of 25 error-recognition and 25 multiple-choice items. They were also asked to report the amount of enjoyment they experienced while answering each type of items. A series of independent samples t-test revealed that male learners performed significantly better on the first type of items while females outscored males in the second type. Moreover, males reported more enjoyment for error-recognition items while the exact opposite was the case for female participants. In other words, the findings revealed that, taking into account the modifications that had to be made to operationalize concepts in this study, Kahneman's statement held ground for male language learners but not females. The bottom line is that this difference was observed across gender and further investigations might shed more light on the reasons behind it. Meanwhile, the results suggest that test designers should put more thought into putting different types of items on grammar tests to avoid tipping the scale to the advantage of one gender.

Keywords: Daniel Kahneman, gender, easier, more enjoyable, error-recognition, multiple-choice

1. Introduction

The fields of study concerned with language teaching, learning, and testing have embraced approaches, theories, and findings from the field of psychology for decades (Mitchell et al., 2019). These sorts of adoptions have actually led to the formation of various successful trends within the practice of language teaching and learning (Ellis, 2015). In recent years, social psychology has risen as another field the findings of which are advocated by language learning practitioners. Social psychology, as a field within humanities and social sciences, is concerned with thought, emotion, and behavior of human beings in social situations (Gilovich et al., 2018). Simply put, utilizing sound theories and assertions from human-related fields is an invaluable asset in the practice of language education. Furthermore, most academic works withing humanities and social sciences have oriented toward interdisciplinary and multidisciplinary approaches due to the complexity of their subject matter i.e., human being (Nowotny et al., 2001; Hvidtfeldt, 2018; Elliott, 2019). Consequently, most recent theories and studies within language related fields are also multidisciplinary in ordination.

Daniel Kahneman (2011), in his seminal work 'thinking, fast and slow', once wrote, "it is much easier, as well as

far more enjoyable to identify and label the mistakes of others than to recognize our own" (Daniel Kahneman, p. 1). The statement clearly elaborates on likely mistakes that are manifested in everyday life. Setting aside the concept of mistake, there are four keywords in what this statement tries to capture regarding human life and mentality: easier, enjoyable, identify (recognize), and label. All these keywords are relevant in the context of language testing in which language learners are often supposed to produce the right answer and/or recognize the existing mistakes in a sentence or text. This study built on the first three keywords to find out whether the statement holds ground in the context of English language learning. Simply put, the ex post facto design of this study aimed at discovering.

The difference between male and female Iranian EFL learners in recognizing grammatical mistakes of their own compared to those of others.

The difference between male and female Iranian EFL learners in the amount of enjoyment they are likely to experience in recognizing grammatical mistakes of their own compared to those of others.

2. Language testing

Simply stated, a language test refers to any procedure that aims at measuring ability, knowledge, and performance on a particular language (Douglas, 2014; Fulcher & Davidson, 2020). The three enumerated concepts are interconnected yet distinct in humanities and social sciences: ability signifies a person's natural potential, knowledge represents the information or understanding that one has acquired through the course of life, and performance refers to actual execution of activities utilizing one's knowledge and abilities (Dweck, 2006). Due to various reasons such as global and economical power, the target language commonly subjected to testing has been English for decades (Crystal, 2003; Jenkins, 2007).

Language tests have been categorized under various labels based on their purpose and format e.g., achievement tests, placement tests, proficiency tests, cloze tests, discrete-point tests, etc. (Brown, 2014; Hughes, 2020). Purpose of a test signifies the reason behind the administration of it as well as what the test intends to measure, while format addresses design, structure, and organization of the test (Bachman & Palmer, 2022). Regardless of the purpose and/or format, tests have always been an inseparable facet of language education. There is an extensive literature on the topic in which language tests have been subjected to scrutiny from various angles (e.g., Coombe, 2012; Fulcher & Davidson, 2013; Lanteigne et al., 2021).

One way through which language tests have been categorized is the availability of predetermined criteria for the scoring procedure. Accordingly, tests are divided into objective tests, in which there exist predetermined criteria so that no judgement befalls the scorer, and subjective tests that require opinionated judgement (Hughes, 2020). Error-Recognition (ER) and Multiple-Choice (MC) are two of the most utilized and well-established types of items in objective tests. In the former, test-takers must indicate the unacceptable part of a sentence from a set of underlined options, while in the latter, they should select the correct answer presented along with multiple distractors (Alderson et al., 1995). Both types of items have been vastly investigated within fields concerned with language learning and teaching in terms of validity, reliability, merits and demerits, etc. (Gorsuch & Griffee, 2017).

The reason behind selecting these two types of items is that ER items resemble identifying others' mistakes, and MC ones could represent awareness of one's own mistakes. It should be noted that the latter is not an exact replacement for awareness of one's own mistakes. However, and taking into account the required objectivity and level of difficulty, MC items were the most logical choice to represent the matter rather than actually utilizing test-takers' actual mistakes in a test. Such approach, given that the purpose is to compare the results to those of ER items, would inevitably invite various threats to validity and reliability of the test as well as the whole research project.

3. Method

This study involved the participation of 150 (75 males and 75 females) Iranian advanced EFL learners between 20 and 30 years of age. To control for the level of English proficiency which could be a threat to ex post facto interpretations, all participants were IELTS 7 holders who were conveniently selected to take part in the study. It should

be noted that all participants consented to take part in this study on account of remaining anonymous. They all took a grammar test consisted of 50 items: 15 ER items followed by 15 MC items, 10 more ER items, and finally 10 more MC items. This format was adopted to control for the fatigue factor that might skew the results if participants were to answer to the whole items of one type and then move on to items of the next type. They all took the test, which lasted 60 minutes, under the same conditions. The items were selected from the grammar section of Ministry of Science, Research, and Technology exam (MSRT) that Ph.D. candidates must take in Iran. This section of the exam targets a wide range of grammatical structures including verbs, nouns, adjectives, adverbs, phrase and clause structures, etc.

The two types of items were scrutinized and judged to be of similar level of difficulty by four language testing experts. At the end of the test, participants were asked to fill out the two-items questionnaire presented in Table 1. They were informed that they should select two numbers between 1 and 10 indicatives of the amount of enjoyment they experienced with regards to ER and MC items, respectively. Therefore, every participant obtained three scores (total score, ER score, and MC score) and two values which were indicative of his/her enjoyment. Finally, a series of independent Samples T-tests were conducted on SPSS 27 to investigate any likely differences across item types, gender, and level of enjoyment.

Table 1. Enjoyment chart

	Extremely Boring											Extremely Enjoyable
Item Type	-	1	2	3	4	5	6	7	8	9	10	-
ER	-	-	-	-	-	-	-	-	-	-	-	-
MC	-	-	-	-	-	-	-	-	-	-	-	-

4. Results

4.1 *Males vs. females in general*

At first, an independent samples t-test was conducted to compare the total score for males and females the results of which, alongside group statistics, are presented in Table 2 and Table 3. There was no significant difference for males (M = 39.8667, SD = 3.82500) and females (M = 40.1467, SD = 3.15669) in the total obtained scores (t (148) = -0.489, p = 0.626, two tailed). The magnitude of the differences in the means (mean difference = -0.28, 95% Cl: -1.41 to 0.85) was minuscule (η^2 = 0.001). The last number is Eta Squared, a measure of effect size, that can range from 0 to 1 and represents the proportion of variance in the dependent variable that is explained by the independent variable (Ary et al., 2018). Simply put, although females performed slightly better, both genders were at the same level of grammatical proficiency regarding the total score.

Table 2. Group statistics for total score across gender

Gender	N	Mean	Std. deviation	Std. error mean
Male	75	39.8667	3.82500	0.44167
Female	75	40.1467	3.15669	0.36450

Table 3. Independent samples t-test for total score across gender

t	df	Sig.	Mean difference	Std. error difference	95% Confidence interval	
					Lower	Upper
-0.489	148	0.626	-0.28000	0.57266	-1.41164	0.85164

4.2 Males vs. females in ER score

The second independent samples t-test targeted ER score for males and females. As presented in Table 4 and Table 5, there was a statistically significant difference for males (M = 21.6667, SD = 2.36719) and females (M = 18.3867, SD = 2.62469) in the obtained ER scores (t (148) = 8.037, p = 0.000, two tailed). Also, the magnitude of the differences in the means (mean difference = 3.28, 95% Cl: 2.47 to 4.08) shows large effect (η^2 = 0.303). In other words, male participants outscored females by a considerable margin in ER items. Data analysis points to the fact that this outperformance across gender is unlikely to be the result of chance factors.

Table 4. Group statistics for ER score across gender

Gender	N	Mean	Std. deviation	Std. error mean
Male	75	21.6667	2.36719	0.27334
Female	75	18.3867	2.62469	0.30307

Table 5. Independent samples t-test for ER score across gender

t	df	Sig.	Mean difference	Std. error difference	95% Confidence interval	
					Lower	Upper
8.037	148	0.000	3.28000	0.40813	2.47349	4.08651

4.3 Males vs. females in MC score

MC score for males and females was the target of the next independent samples t-test the results of which, alongside group statistics, are presented in Table 6 and Table 7. There was a statistically significant difference for males (M = 18.020, SD = 3.11925) and females (M = 21.7733, SD = 1.76737) in the obtained MC scores (t (148) = -8.632, p = 0.000, two tailed). Furthermore, the magnitude of the differences in the means (mean difference = -3.57, 95% Cl: -4.39 to -2.75) resembled a large effect (η^2 = 0.334). To put it into a nutshell, contrary to the results of ER items, female participants outscored males in a meaningful way.

Table 6. Group statistics for MC score across gender

Gender	N	Mean	Std. deviation	Std. error mean
Male	75	18.2000	3.11925	0.36018
Female	75	21.7733	1.76737	0.20408

Table 7. Independent samples t-test for MC score across gender

t	df	Sig.	Mean difference	Std. error difference	95% Confidence interval	
					Lower	Upper
-8.632	148	0.000	-3.57333	0.41398	-4.39140	-2.75526

4.4 *Males vs. females in ER enjoyment*

The next step was comparing males and females in terms of the level of enjoyment they reported regarding ER items. The results of group statistics and independent samples t-test are presented in Table 8 and Table 9. There was a statistically significant difference for males (M = 7.6933, SD = 0.98603) and females (M = 4.2000, SD = 1.09050) in the ER enjoyment level (t (148) = 20.587, p = 0.000, two tailed). The magnitude of the differences in the means (mean difference = 3.49333, 95% Cl: 3.157 to 3.828) was extremely large ($\eta^2 = 0.741$). Differently put, male Iranian EFL learners, regarding ER items, reported more enjoyment of the experience compared to females.

Table 8. Group statistics for ER enjoyment level across gender

Gender	N	Mean	Std. deviation	Std. error mean
Male	75	7.6933	0.98603	0.11386
Female	75	4.2000	1.09050	0.12592

Table 9. Independent samples t-test for ER enjoyment level across gender

t	df	Sig.	Mean difference	Std. error difference	95% Confidence interval	
					Lower	Upper
20.578	148	0.000	3.49333	0.16976	3.15786	3.82880

4.5 *Males vs. females in MC enjoyment*

Finally, the last independent samples t-test aimed at comparing males and females in terms of the level of enjoyment they reported regarding MC items. As presented in Table 10 and Table 11, there was a statistically significant difference for males (M = 4.8800, SD = 1.31478) and females (M = 8.1200, SD = 0.85361) in the MC enjoyment level (t (148) = -17.900, p = 0.000, two tailed). Moreover, the magnitude of the differences in the means (mean difference = -3.24000, 95% Cl: -3.597 to -2.882) was, like that of ER enjoyment, extremely large ($\eta^2 = 0.684$). However, contrary to the result of the previous section, females reported far more experience enjoyment compared to males regarding MC items.

Table 10. Group statistics for MC enjoyment level across gender

Gender	N	Mean	Std. deviation	Std. error mean
Male	75	4.8800	1.31478	0.15182
Female	75	8.1200	0.85361	0.09857

Table 11. Independent samples t-test for MC enjoyment level across gender

t	df	Sig.	Mean difference	Std. error difference	95% Confidence interval	
					Lower	Upper
-17.900	148	0.000	-3.24000	0.18101	-3.59769	-2.88231
-17.900	126.972	0.000	-3.24000	0.18101	-3.59818	-2.88182

5. Discussion and conclusion

The assumption that male and female participants of this study were nearly at the same level of English proficiency is justified as a result of their IELTS score and their total performance in this study. This assumption paves the way for interpreting the findings in light of variables of concern in the current research. The results revealed that the total grammar score of each participant was hardly an amalgam of two similar scores on two types of items. Differently put, participants, more often than not, performed significantly better on one type than the other one. Moreover, whether they performed better on ER items or MC items was significantly impacted by their gender. Simply put, while more percentage of male participants' scores came from ER items, the exact opposite was the case for Iranian female EFL learners. What these findings suggest is that, if transferred to the field of English language learning testing the way it was done in this study, Kahneman's claim, at least among Iranians, is supported among males but not females.

The other main part of his statement, other than it being easier to identify others' mistakes than one's own, was that it actually is more enjoyable. Once again, the same results were observed in the context of Iran: male participants found identifying the mistakes of others more enjoyable but not females. To sum up, Kahneman's statement, taking into account the modifications that had to be made to test his claim, was totally supported for males. Regarding the females, the findings, however, revealed the exact opposite.

Difference across gender is an issue that has been repeatedly supported and discussed withing humanities and social sciences especially in fields such as sociology, psychology, and social psychology (e.g., Coon et al., 2020; Hewstone et al., 2020; Thompson et al., 2016). Gender has been conceptualized as influential in self-construal (e.g., Tanaka, 2023), self-concept (e.g., Wolfram & Gratton, 2014), attributional style (e.g., Hanrahan & Cerin, 2009), happiness (e.g., Stavrova et al., 2012), conformity (e.g., Aronson & Aronson, 2019), aggression (e.g., Bjorkqvist, 2018), motivated behavior (Fathabadi, 2023), etc. The literature on the impact of gender on language learning and achievement also, for the most part, supports the relevance of this factor and its significant impact on various achievement tests (e.g., Pavlenko et al., 2011; Zoghi et al., 2013). However, several studies indicate that gender actually plays a mediating role through other factors such as motivation (e.g., Iwaniec, 2019), beliefs (e.g., Bernat & Lloyd, 2007), and learning styles and strategies (e.g., Viriya & Sapsirin, 2014), etc. The results of this study revealed that the impact of gender on performance of language learners on grammar tests is contingent upon the nature of the items. The implication is that the wiser orientation would be to explore the impact of gender in combination with other factors such as the nature of language test. This way, gender differences are less likely to be averaged out across other factors and thus be overlooked.

The findings of this study should be interpreted with caution, in relation to Kahneman's statement, for three main reasons. Firstly, one might justifiably argue that MC items are not good operational means to measure an individual ability or even tendency to recognize his/her own mistakes. However, the justification behind selecting such items was provided in the earlier parts of this article. Secondly, and in Kahneman's defense, he explained that every author has a setting in mind and his was the office water-cooler. Moreover, the water-cooler that he had in mind, as this information is unknown, might have been surrounded by men! Last but not least, cultural and situational issues should be taken into account since, as various studies suggest, they are influential factors in shaping human behaviors (Aronson et al., 2019; Gilovich et al., 2018; Sadava, 2014).

Regardless of all these issues, his statement held up in reality for males in the context of language learning in Iran with regard to grammar tests. Setting aside Kahneman's statement, the fact is that the difference across gender, in terms of being easier and more enjoyable to recognize mistakes of others, was observed among Iranian EFL learners regarding tests of grammar. This finding is interesting and thought-provoking by itself and could lead up to further investigations. The least one can get from this study is that tests of grammar should include various types of items to avoid tipping the scale into the advantage of one gender over the other.

Conflict of interest

The authors declare no competing financial interest.

References

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge University Press.

Aronson, E., Wilson, T., & Sommers, S. (2019). *Social Psychology* (10th ed.). Pearson.

Ary, D., Jacobs, L. C., Irvine, C. K. S., & Walker, D. (2018). *Introduction to Research in Education*. Cengage Learning.

Bachman, L. F., & Palmer, A. S. (2022). *Language Assessment in Practice: Developing Language Assessments and Justifying Their Use in the Real World*. Oxford University Press.

Bernat, E., & Lloyd, R. (2007). Exploring the gender effect on EFL learners' beliefs about language learning. *Australian Journal of Educational & Developmental Psychology, 7*(1), 79-91.

Bjorkqvist, K. (2018). Gender differences in aggression. *Current Opinion in Psychology, 19*, 39-42.

Brown, J. D. (2014). *Testing in Language Programs: A Comprehensive Guide to English Language Assessment, New Edition*. JD Brown Publishing.

Coombe, C. (2012). *The Cambridge Guide to Second Language Assessment*. Cambridge University Press.

Coon, D., Mitterer, J. O., & Martini, T. S. (2020). *Introduction to Psychology: Gateways to Mind and Behavior*. Cengage Learning.

Crystal, D. (2003). *English as a Global Language*. Cambridge University Press.

Douglas, D. (2014). *Understanding Language Testing*. Taylor & Francis.

Dweck, C. S. (2006). *Mindset: The New Psychology of Success*. Random House.

Elliott, A. (2019). *Critical Visions: New Directions in Social Theory*. Routledge.

Ellis, R. (2015). *Understanding Second Language Acquisition*. Oxford University Press.

Fathabadi, J. (2023). The hidden side of learning English as a socially motivated behavior in Iran educational system. *Interchange, 54*(3), 155-171.

Fulcher, G., & Davidson, F. (2020). *Language Testing and Assessment: An Advanced Resource Book*. Routledge.

Fulcher, G., & Davidson, F. (2013). *The Routledge Handbook of Language Testing*. Routledge.

Gilovich, T., Keltner, D., Chen, S., & Nisbett, R. E. (2018). *Social Psychology*. WW Norton & Company.

Gorsuch, G., & Griffee, D. (2017). *Second Language Testing for Student Evaluation and Classroom Research*. IAP.

Hanrahan, S. J., & Cerin, E. (2009). Gender, level of participation, and type of sport: Differences in achievement goal orientation and attributional style. *Journal of Science and Medicine in Sport, 12*(4), 508-512.

Hewstone, M., Stroebe, W., & Jonas, K. (2020). *An Introduction to Social Psychology*. Wiley.

Hughes, A. (2020). *Testing for Language Teachers* (2nd ed.). Cambridge University Press.

Hvidtfeldt, R. (2018). *The Structure of Interdisciplinary Science*. Palgrave Macmillan.

Iwaniec, J. (2019). Language learning motivation and gender: The case of Poland. *International Journal of Applied Linguistics, 29*(1), 130-143.

Jenkins, J. (2007). *English as a Lingua Franca: Attitude and Identity*. Oxford University Press.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Penguin Books.

Lanteigne, B., Coombe, C., & Brown, J. D. (2021). *Challenges in Language Testing Around the World*. Springer.

Mitchell, R., Myles, F., & Marsden, E. (2019). *Second Language Learning Theories*. Routledge.

Nowotny, H., Scott, P., & Gibbons, M. (2001). *Re-thinking Science: Knowledge and the Public in an Age of Uncertainty*. Polity Press.

Pavlenko, A., Blackledge, A., Piller, I., & Teutsch-Dwyer, M. (2011). *Multilingualism, Second Language Learning, and Gender* (Vol. 6). Walter de Gruyter.

Sadava, S., & Alcock, J. (2014). *An Introduction to Social Psychology: Global Perspectives*.

Stavrova, O., Fetchenhauer, D., & Schlösser, T. (2012). Cohabitation, gender, and happiness: A cross-cultural study in thirty countries. *Journal of Cross-Cultural Psychology, 43*(7), 1063-1081.

Tanaka, M. (2023). Motivation, self-construal, and gender in project-based learning. *Innovation in Language Learning and Teaching, 17*(2), 306-320.

Thompson, W. E., Hickey, J. V., & Thompson, M. L. (2016). *Society in Focus: An Introduction to Sociology*. Rowman & Littlefield Publishers.

Viriya, C., & Sapsirin, S. (2014). Gender differences in language learning style and language learning strategies. *Indonesian Journal of Applied Linguistics, 3*(2), 77-88.

Wolfram, H. J., & Gratton, L. (2014). Gender role self-concept, categorical gender, and transactional-transformational leadership: Implications for perceived workgroup performance. *Journal of Leadership & Organizational Studies, 21*(4), 338-353.

Zoghi, M., Kazemi, S. A., & Kalani, A. (2013). The effect of gender on language learning. *Journal of Novel Applied Sciences, 2*(4), 1124-1128.