Volume 3, Issue 1, Spring-Summer 2019 (Biannual – Serial No. 5)

Developing and Validating a New Version of an EFL Multiple-Choice Reading Comprehension Test Based on Fuzzy Logic[1]

Amir Hossein Shahballa[2]
Behzad Ghonsooly*[3]
Hosein Karami[4]

Received: 2019-07-05 | Revised: 2019-09-14 | Accepted: 2019-09-16

Abstract

Multiple-choice tests do not assess examinees' knowledge in accord with reality. In fact, the partial knowledge of examinees is not assessed. Providing a new approach to the assessment of reading comprehension in the framework of fuzzy logic, this study aims to measure this partial knowledge. In this approach, participants have to choose as many correct options as there are considering the stem. Therefore, the correct answer to each question can range from one option to all options. For the first

[1] DOI: **10.22051/lghor.2019.27083.1154**
[2] PhD Candidate, Department of English, University of Tehran, Kish International Campus; ashahballa@ut.ac.ir
[3] Professor of Applied Linguistics, Department of English Language and Literature, Ferdowsi University of Mashhad, (Corresponding author);ghonsooly@um.ac.ir
[4] Assistant Professor of English Department, University of Tehran; hkarami@ut.ac.ir

session, an expository and an argumentative genre, and for the second session, the reading section of a TOEFL test was used. The results showed that the approach is fairer as it considers the partial knowledge of the examinees while in other common multiple-choice tests this is ignored. Also, the use of idea units as the units comprising a text gives us clues regarding the degree of difficulty of different parts of a text as well as clues about why misunderstanding may occur of the same text among different people.

Keywords: Reading comprehension, Fuzzy logic, Multiple-choice tests, Partial knowledge, Fair assessment

Introduction

In the real world, there exist different problems which, according to Harris (2000), are one of the three categories. First, in many areas of study such as engineering, quantitative problems are solved in a deterministic way. In other words, they are solved using classical (Aristotelian) logic which uses sets with sharp boundaries. It means that something either is or is not a member of the sets. Second, statistical methods are applied to account for the random uncertainty inherent in some problems. Here, the concept of probability is used. And third, in many real problems, the uncertainty is neither random nor of a statistical nature. Here, the corresponding concept is possibility rather than probability. It is for this last category which fuzzy logic (FL) is useful.

According to Zadeh (2008), the inventor of fuzzy logic, "There are many misconceptions about fuzzy logic. Fuzzy logic is not fuzzy. Basically, fuzzy logic is a precise logic of imprecision and approximate reasoning" (p. 2751).

An example by McNeill and Thro (1994) clarifies how fuzzy logic can be a precise logic, using a bowl of oranges! Is the following figure a bowl of oranges? The answer is no.

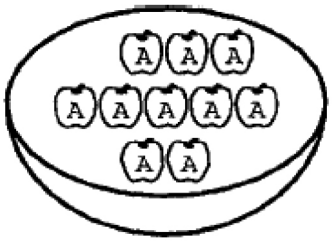

Figure 1. A bowl of apples

What about the next figure? Is it a bowl of oranges?

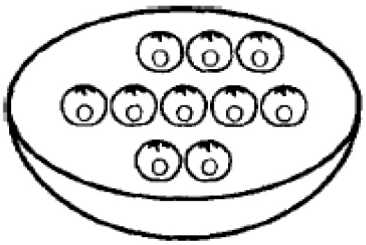

Figure 2. A bowl of oranges

The answer is yes. As McNeill and Thro (1994) say, this is an example of crisp logic. The bowl contains *either* totally apples *or* totally oranges. However, as they accurately say, life is more complex. Consider for example the following figure in which someone has swapped one orange for one apple in the apples bowl. Is it a bowl of oranges?

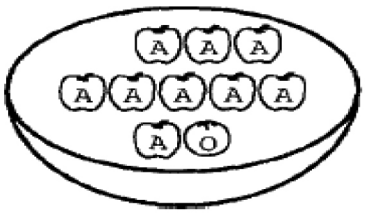

Figure 3. A bowl of apples and oranges

Here, as they say:

> The situation itself makes *either* Yes *or* No inappropriate. In fact, if you had to say Yes or No, your answer would be less precise than if you answered One, or Some, or A Few, or Mostly—all of which are fuzzy answers, somewhere in between Yes and No. They handle the actual ambiguity in descriptions or presentations of reality. (p. 6)

The following figure shows the range of possible situations and answers to "Is it a bowl of oranges?"

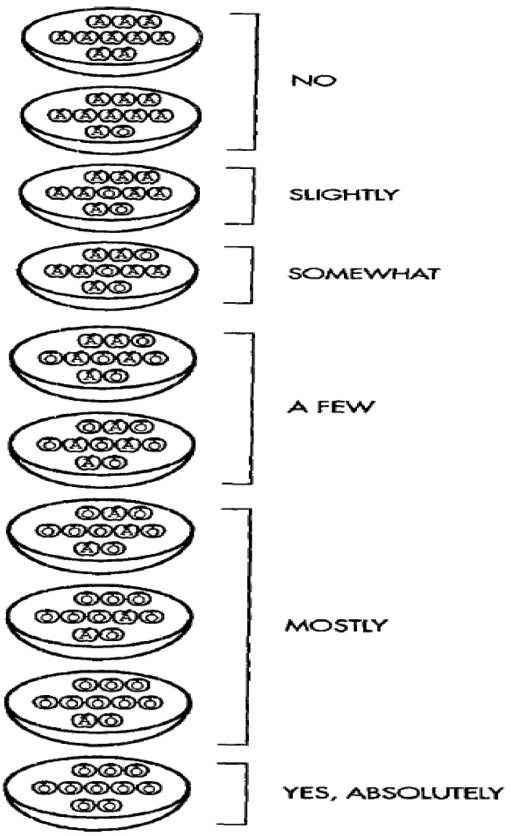

Figure 4. The range of answers taken from McNeill and Thro (1994, pp. 3-5)

Imprecision is, in fact, a characteristic of our understanding of the world phenomena, including understanding of texts. When individuals read a piece of text, we cannot say they have precisely grasped what the text conveys; nor can we say they have understood nothing. Rather we can say their understanding is poor, not bad, well, fairly well, etc.

Literature Review

Over the years, assessment has taken many forms, including multiple-choice, true/false, and essay type questions. However, the one which is most commonly used is the multiple-choice form. This is because of the advantages this form offers both in administering and in scoring. However, the way multiple choice tests assess examinees' knowledge is not in accord with reality, i.e. They consider an examinee's answer to an item to be either totally correct or totally wrong. This is the case even when the best choice is required. In fact, multiple-choice questions do not consider the partial knowledge of examinees. In other

words, examinees may know part of the answer; however, as there is no possibility for expressing their knowledge through the options provided, they do not gain any credit. Considering this limitation, some attempts have been made to devise questions based on fuzzy logic.

One attempt is Faghih and Alamdar's (2007) study which introduced a new form of answer sheet (designed for Persian questions, i.e. questions from right to left):

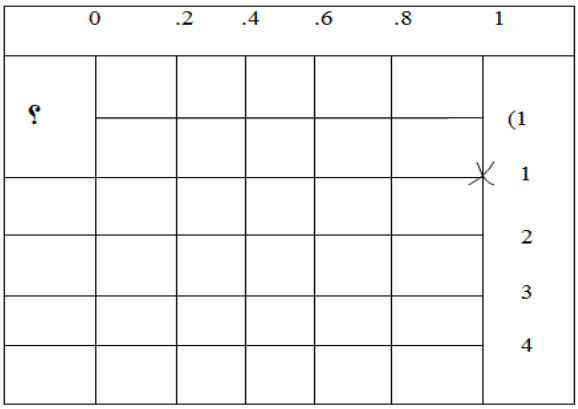

Figure 5. An example of fuzzy multiple-choice answer sheet

In this model, the examinees can choose the degree of confidence in the option they select. If Option 2 with the degree .7 is marked, it means that the examinee is sure to the degree .7 or (70%) that the correct option is Option 2. However, the examinee considers the selection to be wrong to the degree .3. Therefore, since the correct option is 1 and the examinee has pointed out his/her uncertainty, the score given to the answer is .3.

Another attempt is the partial-knowledge-based answer sheet which Zahedi (2001) introduced:

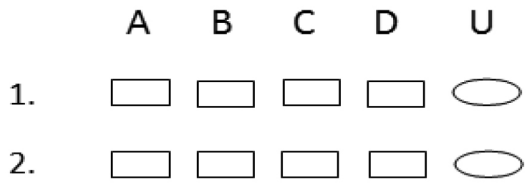

Figure 6. An example of fuzzy multiple-choice answer sheet

In the proposed answer sheet, examinees can indicate their uncertainty by marking the U(uncertainty) option.

Zahedi (2001) mentioned that partial knowledge (or, in his terms, parledge) can capture the following five response modes:

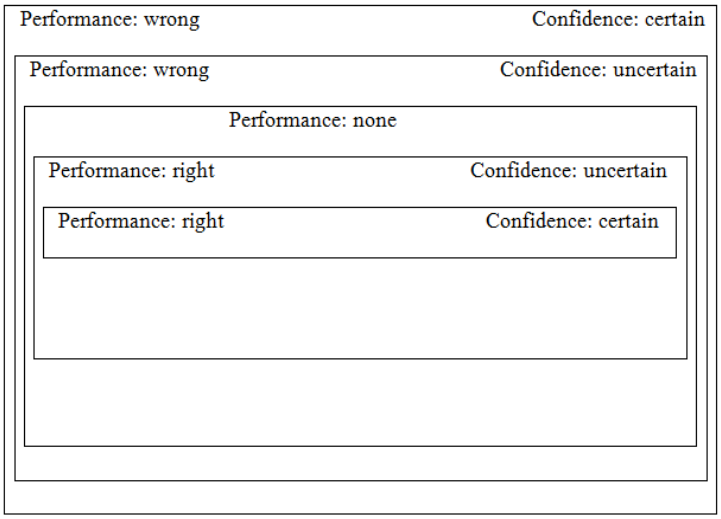

Figure 7. Fuzzy response modes for parledge

In his model, the parledge scoring procedure is as follows:

Performance confidence score
correct certain 1
correct uncertain 3/4
no-performance
incorrect uncertain -1/4
incorrect certain -1/3

Figure 8. Numerical values for a four-choice multiple-choice test

Another study is the one by Cin and Baba (2008) who, in order to evaluate the English proficiency of students, developed a fuzzy multi-criteria assessment software program which was multiple-purpose and user friendly.

Also, Baba, Bakanay, and Cin (2012) used fuzzy logic in their study. They developed a fuzzy system for evaluating students' projects in engineering education. Another study using fuzzy logic for assessment is that by Baba, Kuscu, and Han (2009). They developed a fuzzy multi-criteria decision-making software program which enables the user to make evaluations according to the number of decision-makers and evaluation weights of criteria. Another study is Al-Hammadi and Milne's (2004) which developed a neuro-fuzzy model to evaluate

and predict the performances of students before admitting to the college based on their secondary-school and entry exam results. Shahballa and Alamdar Youli (2012) is another study which developed a fuzzy system which had two inputs, i.e. two scores: one, the main idea (MI) score and another, the idea unit (IU) score. These scores were gained after scoring the examinees' understanding. In fact, they wrote their comprehension of the text and then their comprehension was scored. Then, the two scores were fed into a fuzzy system and a fuzzy score was yielded. In fact, "Fuzzy set theory is an efficient and effective method to represent the uncertainty and fuzzy terms in the assessment environments" (Al-Hammadi & Milne, 2004, p. 837).

Keeping the advantages of multiple-choice tests, this study attempted to eliminate the disadvantage of the binary way of looking at answers as right or wrong and provide a fairer approach to assessment.

According to Cohen, Manion, and Morrison (2000):

> Fuzzy logic enables us to gain a more precise measurement of the variance *within* and *between* these semantic categories; it recognizes that imprecision, rather than bivalence (*either* something is *or* is not the case) is a characteristic of many phenomena. Fuzzy logic opts for shades of greyness rather than black-or-white (Kosko, 1994:102)! In the field of education Fourali (1997) has shown how fuzzy logic is particularly useful in assessment. (p. 389)

To gain scores based on fuzzy logic, a unit of measurement was needed. A review of the literature revealed that different studies have used different units for analysis. One unit of analysis is the t-unit (Andoline, 1980; Budd, 1988; Distefano & Valencia, 1980; Hosseinchary & Yosefy, 2008; Samim-Banihashemi, 1992). Bardovi-Harlig and Bofman (1988) and Bardovi-Harlig (1992) consider sentence a better unit for analysis than t-unit. Roll, Frid, and Horne (2007) developed and used a tool named "GRAMMAL". Also, Rashidi and Shahballa (2010) and Shahballa and Alamdar Youli (2012) used IUs as a unit of measurement.

This study utilized IUs as the basic elements of scoring in a fuzzy framework. Chafe (1985) defines IUs as "clearly identifiable elements of spoken language" (p. 106). According to Gee (1996), the "vast majority of these idea units are a single clause, with one piece of new information towards the end of the clause. It is only when the subject of the clause, or an adverbial element, is new information that it constitutes an idea unit by itself" (p. 103).

Idea units in this study are exactly equal to propositions. Richards (1983) defines propositions as "the basic units of meaning involved in comprehension" (p. 220).

As an example, a common multiple-choice question with Option A as the correct answer and a fuzzy multiple-choice question with Options A and B as the correct answer are compared. The former as the representative of a binary framework and the latter as the representative of a fuzzy framework are differ-

ent in that the former treats the response to the question as either completely true or completely false. This has two drawbacks. In the first place, if the response is correct, it may be a combination of the test-taker's knowledge and chance. There is a chance of 25% to choose the correct answer. In fact, the correct answer can be only Option A, only Option B, only Option C, or only Option D. In the second place, if the response is wrong while the test-taker knows part of the answer, the question has not considered the partial knowledge of the test-taker. This seems unfair. However, regarding the fuzzy question, in the first place, if the test-taker chooses only Option A, or only Option B, he/she will gain scores for his/her knowledge. Second, as the test-taker has fifteen possibilities to choose the correct answer, the probability of choosing an option based on chance is about 6.66% and not 25%. In fact, the correct answer can be only Option A, only Option B, only Option C, only Option D, Options A and B, Options A and C, Options A and D, Options B and C, Options B and D, Options C and D, Options A, B and C, Options A, B and D, Options A, C and D, Options B, C and D, or Options A, B, C and D.

The research questions are as follows:

1. Is the new version of EFL multiple-choice reading comprehension test which is based on fuzzy logic valid?
2. Is there a significant difference between participants' understandings in testing through common multiple-choice questions (of which TOEFL reading comprehension scores are a representative) and through fuzzy multiple-choice questions?
3. What are the test-taking strategies participants use when answering fuzzy questions?

Method

Participants

Participants in this study included 200 male and female Iranian university students and English language learners. In fact, 150 of the students were from different universities, the majority of them being from the University of Kashan. To ensure an appropriate level of proficiency, some classes were chosen after consulting the head of the English department or the teacher. The days and times for giving the test were arranged with the related teachers. The test was administered to the students in the second half of the class or in the last half an hour of the class. The teacher encouraged the students to participate in the test. Also, as an incentive, a discount of 40 to 60% to buy English books from a main publisher of English material in Iran was considered for those participating. As a result, those consenting to participate in the study took the test and some left the class. The other participants were from different institutes and took the test voluntarily. Nearly all the participants were third-year students or at higher levels and their level of language proficiency was intermediate and above. However, some of them did not answer the TOEFL test questions completely. Therefore, out of 200 answer sheets, only 136 were useable for the study, seven

of which were deleted after using Rasch analysis on the TOEFL section of the test. As a result, 129 answer sheets were analyzed for the purposes of this study.

Materials

Two text types, an expository and an argumentative text, were selected from Kirszner and Mandell (2012). These two text types were chosen since these are the types mostly used in academic course books and students mostly deal with them in their education. Mariconda (2001) mentioned that expository texts - "texts written for the purpose of informing others" (p. 13) - play a major role in middle grades where students must apply reading skills to obtain critical information. She also states: "In fact, most of the writing we do as adults is expository. Taking notes, jotting phone messages, composing business letters- these are all ways we convey information in writing." (p. 13). Andrews (2010) mentioned some reasons regarding the importance of argument in higher education. These include the expectation of students to argue rationally in higher education, the advancement of knowledge through argument, argument being related to clarifying and persuading, and argument being encouraged at universities as it is enjoyable. These reasons indicate the expectation and encouragement of students to use argumentation in higher education.

Questions 1-19 were developed for the expository text (containing 632 words) and Questions 20-42 were developed for the argumentative text (containing 653 words). Also, the reading section of a TOEFL test (Form 0401 Administered by ETS in 2004) was used which contained five passages (1759 words) and 50 questions in total.

Procedure

The participants, in the first session, received two reading comprehension texts, an expository and an argumentative text, with multiple-choice questions developed based on fuzzy logic. For each question, five options had been developed. Each option was in fact an IU. The participants were supposed to check as many correct options as possible on the answer sheet. The correct response could range from 1 option correct to all options correct. In the second session, the reading section of a TOEFL test was given to the participants. For both fuzzy multiple-choice questions and TOEFL questions (used as representative of common multiple-choice questions), negative marks were considered.

As already mentioned, five options were developed for each question in this study and, therefore, there could exist five different states. In other words, the correct answer for each question could be one option,two options, three options, four options, or five options. Regarding the first state, i.e. having only one option as the correct answer, there were five possibilities. The test-taker could choose 1) only Option A, 2) only B, 3) only C, 4) only D, or 5) only E. As to the second state, i.e. having two options as the correct answer, there were 10 pos-

sibilities. The correct answer for each question could be 1) Options A and B, 2) A and C, 3) A and D, 4) A and E, 5) B and C, 6) B and D, 7) B and E, 8) C and D, 9) C and E, or 10) D and E. Regarding the third state, i.e. having three options as the correct answer, there were nine possibilities. The correct answer could be 1) A, B, C, 2) A, B, D, 3) A, B, E, 4) A, C, D, 5) A, C, E, 6) A, D, E, 7) B, C, D, 8) B, C, E, or 9) C, D, E. Regarding the fourth state, i.e. having four options as the correct answer, there were three possibilities: 1) A, B, C, D, 2) A, B, C, E, or 3) B, C, D, E. Finally, regarding the fifth state, i.e. having five options as the correct answer, there was only one possibility which was choosing all five options.

Such a fuzzy question has some advantages. On the one hand, contrary to common multiple-choice questions where the test-takers have to choose only one option, choosing one option in such a fuzzy multiple-choice question does not prevent the test-takers from choosing other options. This helps them to gain scores for their true selections. On the other hand, the different number of possibilities can decrease answering by chance.

The questions for the expository text consisted of one question with three options as the correct answer, asking for the MI of the text, and 18 questions for the IU part of the questions, mainly dealing with the surface understanding of the text. These 18 questions comprised one question with one option as the correct answer, three questions with two options as the correct answer, seven questions with three options as the correct answer, six questions with four options as the correct answer, and one question with five options as the correct answer.

The questions for the argumentative text consisted of one question with three options as the correct answer, asking for the MI of the text, and 22 questions for the IU part of the questions, mainly dealing with the surface understanding of the text. These 22 questions comprised two questions with one option as the correct answer, right questions with two options as the correct answer, eight questions with three options as the correct answer, two questions with four options as the correct answer, and two questions with five options as the correct answer.

To consider guessing and the related penalties, the formula R-[w/ (n-1)](R, the number of correct answers, W, the number of wrong answers, and n, the number of options) was applied.

To examine the test-taking strategies examinees used while answering fuzzy questions, six of the examinees were asked to think aloud their test-taking strategies while answering both the expository text and the argumentative text. Four of the verbal reports were suitable for analysis. As Ghonsooly (2013) mentioned, rarely does the number of participants in the think-aloud investigations exceed 10 at the level of Ph.D. To exemplify, he refers to Cavalcanti (1983) using four Spanish participants, Block (1986) asking nine students to think aloud, and Sarig (1987) including 10 participants for the think-aloud procedure.

For the think-aloud part of the study, an example text and some example questions were given to the participants and the procedure was clarified to

them. Then, the participants were asked to provide the researcher with what goes on in their mind while trying to answer the fuzzy questions. In this way, four participants provided the researcher with eight recordings of both the expository and argumentative texts. Then, the recordings were transcribed and the coding procedure started which used the followings abbreviations: TR (text reading), SR (stem reading), OR (option reading), OS (option selection), TREP (text reprocessing), OT (option translation), SDQ (self-directed question), TT (text translation and trying to understand the text), MKWS (marking key words in the stem), MKWT (marking key words in the text).

The coding procedure has been displayed using part of the verbal report of Participant 1 in the following. The underlined parts were in Persian and then translated into English for the sake of reporting here. The abbreviations show what strategies have been used.

(reading the instruction and the text aloud until 5:30, TR1, then reading the stem of question 1, SR1 and each option OR1 until 05:55) Well, it says the MI of the text ST1, is not an important issue (option c), OR2, it was important, so Option C is not correct, 06:04, Creates a negative situation (Option B), This may be correct, Option B is correct, leads to overlooking violence against women (Option A) No, it neglected violence against women, it's not correct, OT1, so Option B, negative idea is true till now 06:22, Encourages violence against women (Option D), In my opinion, B and D are true, it increases violence against women, OT2, is only (Option D), so I mark B and D, in my opinion, B and D are true, OS1**,** 06:35.

(reading the stem of question 2 aloud, SR2, and options, OR3, until 06:58), Well, which of these are true?, SDR1, Women wore these, OR4, OT3, I don't know, I think it wasn't so (Option E), the T-shirts reappeared again in (Option D), I'm not sure about this year, I have to return to see if it is there or not, TREP1, again 1950, their previous greatest resurgence occurred in the 1950s, this seems to be true, The T-shirt interests only the youth (Option C), it is ... only for young people, no, I think it was also for people aged 25, 30, 31, the T-shirts can't be worn alone (Option B), I don't know this, some fashion designers have decorated (Option A), I think A and D are true, OS3, 07:41.

(reading the stem of question 3 aloud, SR3), well, Paragraph 1, Line 3, TREP 2, (returning to text), the previous greatest resurgence…, OR5, "one" refers to suit (Option E), fashion (Option D), jewels (Option A) this is not true, wife-beater (Option B) this is not true, what were they wearing? SDQ2, (reading the text aloud again) TREP3, it refers to wife-beater, 08:32, "fashion" is not true, "suit" is not true, "Gap" is also, I thinks it is also "jewels", A and B can be true, OS4, 08:46.

Results

In the version of multiple-choice questions proposed in this study, the examinees need to mark as many correct options as possible in relation to the item. In

this approach, the two inputs gained from a single fuzzy test provide us with both a deep understanding and a surface understanding of the text. In other words, the MI feeds the deep understanding to the fuzzy system and the IU feeds the surface understanding to the fuzzy system. Therefore, the final score resulting from the fuzzy system is a combination of deep and surface understating. The structure of the fuzzy system developed by Shahballa and Alamdar Youli (2012) is as follows:

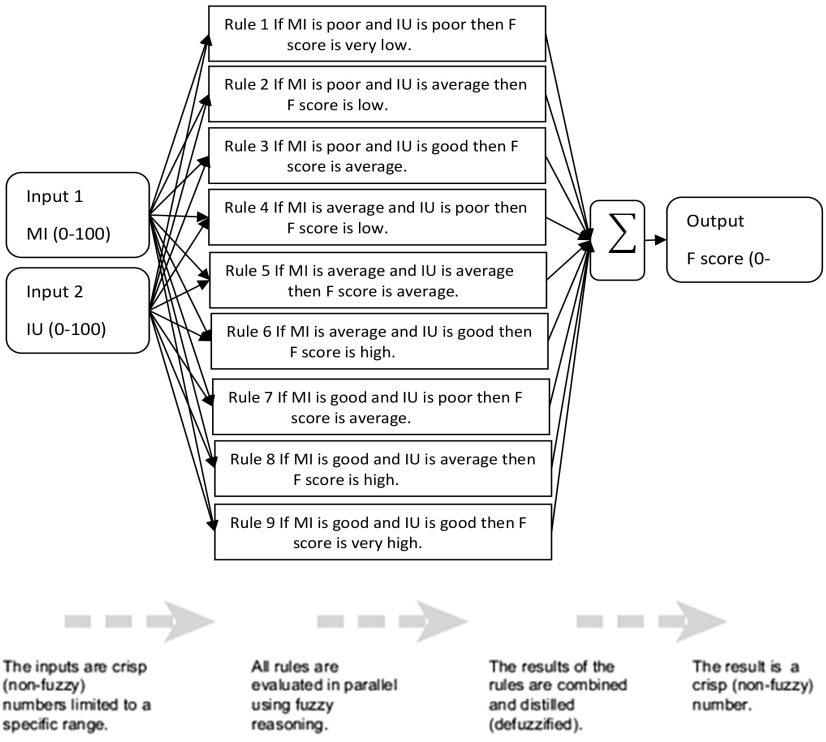

Figure 9. The structure of the fuzzy system

The scores for the TOEFL section were calculated through the formula R-[w/(n-1)] and then the percentage was the final score. In the fuzzy section, for each text, first the formula R-[w/(n-1)] was used for the first question of the text dealing with the MI of the text and it was also used for other questions dealing with IUs; then, the percentages of the two scores (MI and IU) were gained and fed into the fuzzy system developed by Shahballa and Alamdar Youli (2012) and a fuzzy score was obtained. For the entire fuzzy section, the average of the percentages of the MI scores of the two texts and the average of the percentages of the IU scores of the two texts were fed into the fuzzy system, and a fuzzy score for the entire fuzzy section was gained. The binary scores, i.e. the TOEFL scores and the fuzzy scores, are tabulated in the following. A comparison of fuzzy scores and TOEFL scores shows the consideration of partial

knowledge in the test-takers' answers as evident in the following table. Except for 42 of the participants (in bold in Table 1) whose fuzzy scores were smaller than their TOEFL scores (common scores), other participants, i.e. about 70% of them, had fuzzy scores greater than their TOEFL scores. As the fuzzy scores are higher, this suggests the possibility in the test for participants to express their partial knowledge.

Table 1.
Fuzzy and TOEFL Scores

Partici tici- pant No.	F scores	TOEFL	Partici tici- pant No.	F scores	TOEFL	Partici tici- pant No.	F scores	TOEFL
201	45.5	30.66	246	11.5	8.66	**289**	**8.67**	**41.33**
202	74.4	25.33	247	75.2	46.66	**290**	**29.3**	**62.66**
203	78.2	46	**248**	**8.67**	**12.66**	291	50.9	30.66
204	51.7	23.33	**249**	**26.4**	**38**	292	60	47.33
205	**25.7**	**30.66**	250	44.9	2.66	***293***	***12.6***	***13.33***
206	**17.9**	**22.66**	251	54.4	17.33	**294**	**24.7**	**46**
207	55	9.33	252	17.4	16	295	52.6	35.33
208	22.5	1.33	253	28.6	23.33	**296**	**34.5**	**56**
209	**8.67**	**38**	254	56.1	10	297	64.2	5.33
210	67.8	37.33	255	33.5	8.66	298	66.7	16
211	***35.4***	***37.33***	256	91.3	26	**299**	**24.3**	**25.33**
212	**8.67**	**20**	257	60	25.33	300	67	48
213	72.5	38.66	258	42	6.66	301	51.7	38.66
214	**23.6**	**52**	259	21.5	7.33	**302**	**32.5**	**44**
215	**24.3**	**32**	260	25.9	44.66	303	76.4	58
216	19.9	0	**261**	**8.67**	**60**	***304***	***91.3***	***92***
217	12.6	4	262	53.8	46	306	64.9	49.33
218	77.4	50.66	263	37.8	1.33	308	17.9	0
219	**15.9**	**36**	**264**	**8.67**	**18**	309	64.7	46
220	23.7	12	265	60.9	0	310	73.3	22.66
221	37.7	26.66	266	51.4	2.66	**311**	**78.7**	**86.66**
222	21.9	12	267	27.5	33.33	312	91.3	84
223	30.5	22.66	268	26.8	0	313	69.4	65.33
224	21.8	21.33	269	34.2	12.66	314	63.3	49.33
225	20.4	11.33	270	22.1	12.66	316	28.6	20.66
226	54.1	30	271	43.7	31.33	317	35.3	34.66
228	23.7	12.66	272	25.7	0	318	45.6	20
229	8.67	0.66	273	74.7	30	319	38.3	17.33
230	23.6	10	274	45.3	35.33	320	41	10
231	25.4	13.33	**275**	**14**	**31.33**	322	46.1	18
232	80.2	53.33	276	74.4	32	***324***	***43.7***	***44***
233	8.67	6.66	**277**	**8.67**	**25.33**	325	45.9	4
234	**11.8**	**31.33**	278	55.2	28.66	326	91.3	78.66
235	**36.1**	**46.66**	279	46.9	21.33	327	82.1	33.33
236	**32.3**	**60.66**	280	61.1	20	**328**	**28**	**64.66**
237	**30.3**	**52.66**	**281**	**25.7**	**30.66**	**329**	**44.6**	**68**
238	41.2	0	282	50.4	20	**330**	**45.4**	**56**
239	61.3	18	**283**	**8.67**	**24**	**331**	**71.1**	**86.66**
240	18.1	18	284	49.5	33.33	332	51.4	20
241	33.1	6.66	**285**	**56.6**	**60.66**	**333**	**31.8**	**65.33**
242	**8.67**	**11.33**	286	74.7	34.66	**334**	**28.1**	**68**
243	12.6	2.66	**287**	**40.6**	**63.33**	**335**	**74.3**	**89.33**
245	**8.67**	**19.33**	**288**	**8.67**	**18**	336	26.1	24.66

The research questions are dealt with in the following way:

1. Is the new version of EFL multiple-choice reading comprehension test which is based on fuzzy logic valid?

To check the concurrent validity of the test developed based on fuzzy logic, a Pearson correlation test was run between the TOEFL test and the test developed based on fuzzy logic.

Table 2.
Correlation between the Fuzzy Test and the TOEFL Test

		TOEFL	F score total
TOEFL	Pearson Correlation	1	.432**
	Sig. (2-tailed)		.000
	N	129	129
F score total	Pearson Correlation	.432**	1
	Sig. (2-tailed)	.000	
	N	129	129

As indicated in Table 2, the correlation coefficient was .43, significant at .05 level.

2. Is there a significant difference between participants' understandings in testing through common multiple-choice questions and through fuzzy multiple-choice questions?

To see whether the participants' understandings differed in testing through common multiple-choice questions and through fuzzy multiple-choice questions, a paired samples t-test was run between fuzzy scores and scores given to common multiple-choice questions.

Table 2.
Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	F score total – TOEFL	1.17004E1	23.88381	2.37653	6.98543	16.41536	4.923	100	.000

The difference is significant at .05 and, as shown in Table 1, the participants gained better scores in the fuzzy section.

The following figure demonstrates the difference in fuzzy and TOEFL scores and the fact that fuzzy scores are greater than TOEFL scores.

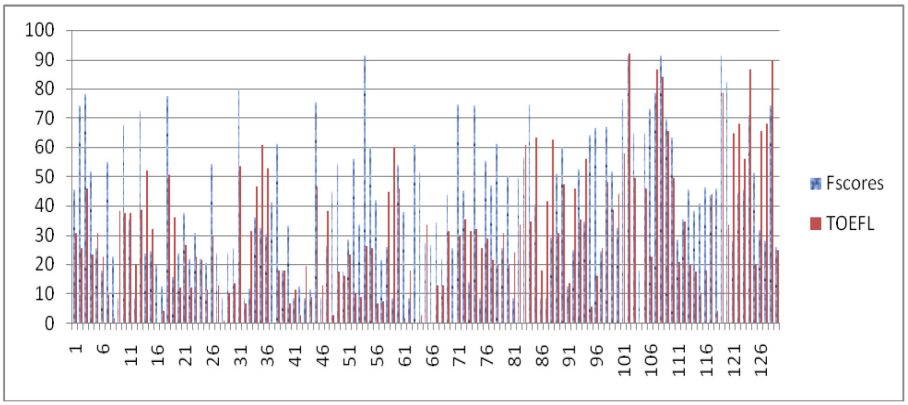

Figure 10. Fuzzy scores and TOEFL scores for each participant

3. What are the test-taking strategies participants use when answering fuzzy questions?

For the think-aloud part of the study, an example text and some example questions were given to the participants and the procedure was clarified to them. Then, they were asked to provide the researcher with what goes on in their mind while trying to answer the fuzzy questions. This way, four participants provided the researcher with eight recordings of both the expository and argumentative texts. Then, the recordings were transcribed and the coding procedure started which used the followings abbreviations: TR (text reading), SR (stem reading), OR (option reading), OS (option selection), TREP (text reprocessing), OT (option translation), SDQ (self-directed question), TT (text translation and trying to understand the text), MKWS (marking key words in the stem), and MKWT (marking key words in the text).

To show the following step, downward arrows were used in the figures. Also, parentheses indicate that the strategy occurred at some points, not all the time.

Regarding the expository text, on the whole, Participant 1 first started to read the whole text. Then, he started to read the stem and the options of the question. Where necessary, the participant returned to the text and read the related part. Also, where necessary, he translated the option to understand the meaning and, at some parts, he translated the text while returning to and reading it. Finally, he chose the correct options. Participant 2, after reading the first sentence of the text and reading the stem and options of Question 1, tried to find the MI of the text. However, the general pattern she used to answer the questions was first reading the stem of the question and using a self-directed question at some points, then reading the options and next returning to the text, again reading the options and finally choosing correct options. Participant 3 first read the stems and marked the keywords in stems and then read the text

to find and mark the keywords marked in the stems. Then, she started to read the stems and options and returned to the text to find the answer and finally selected the correct options. Participant 4 started to read the text and translate some parts. After that, she read the stem and the options and where necessary she used the strategy of translating the option. Moreover, at some parts, she used self-directed questions. Finally, she either chose the correct options and then returned to the text to check the answer or read the text to find the answer and then chose the correct options.

The following patterns or trends were obtained after analyzing the transcripts related to the expository text.

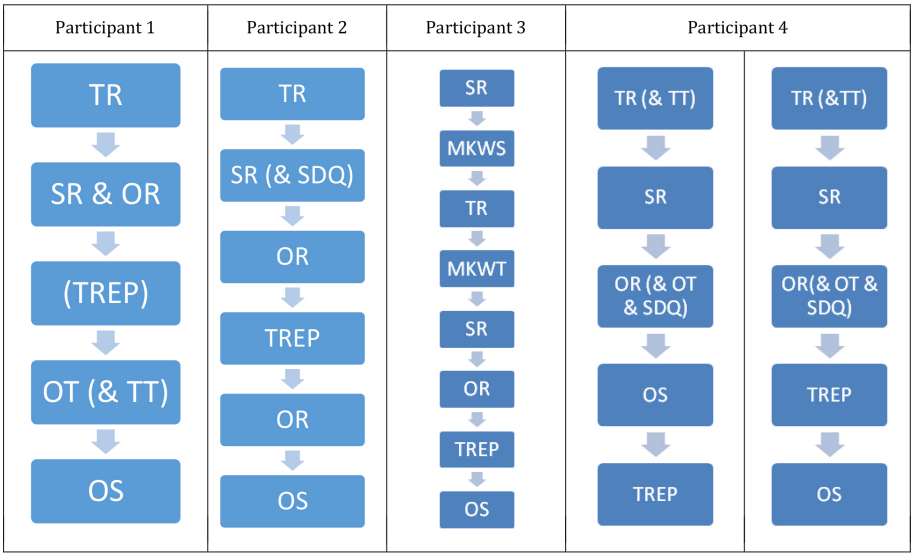

Figure 11. Think-aloud patterns for the expository text

Regarding the argumentative part, the first participant started to read the text and translated the text after reading each paragraph. After finishing reading, he started to read the stem and, where necessary, returned to the text. Also, at some points, he used self-directed questions. After that, he read the options. Again, at some points, he referred to the text to find the answer. Finally, he selected the correct options. The second participant first read the text. Then, after reading the stem and options, she chose the correct options. Then, she returned to the text to check whether the answers were correct or not. Participant 3 first read the questions quickly. Then, she read the text and at some points she translated some parts to understand it better. After that, she read the stem and at some points translated the stem. Next, she returned to the text. After reading options and at some points returning to the texts again, she chose the correct option. Participant 4 read part of the text and then read several stems. After

reading some stems, she translated them. Then, she continued reading the text. After finishing reading, the pattern she used for answering was first reading the stem and at some points translating it. After that, either she read the options and returned to the text or vice versa she returned to the text and then she read the options. Finally, she chose the correct options.

The following patterns or trends are related to the argumentative text.

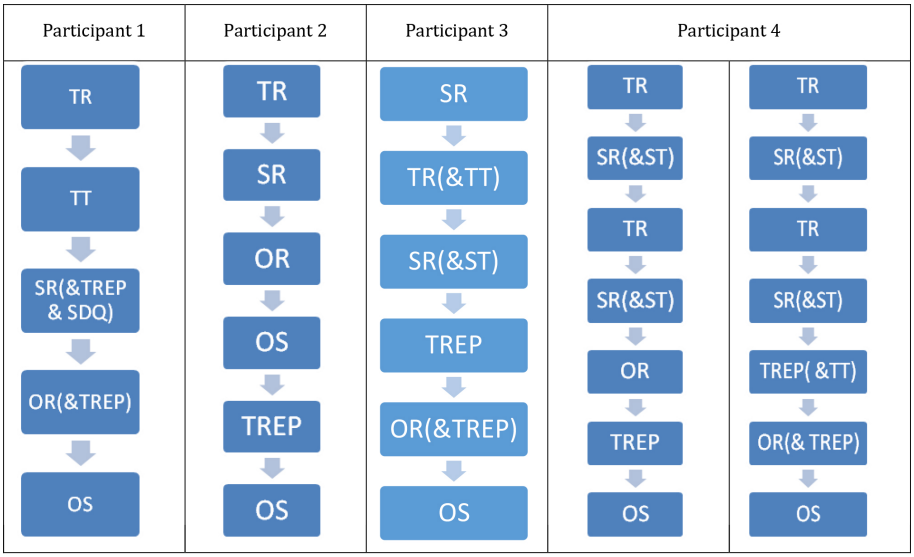

Figure 12. Think-aloud patterns for the argumentative text

The following tables indicate the frequency of the strategies used by participants.

Table 3.
Frequency of Strategies in the Expository Genre

Strategy	Participant 1	Participant 2	Participant 3	Participant 4
TT	1			
SR	19	20	32	22
ST		5		3
SDQ	3	1		3
OR	23	19	13	22
OT	4	3		11
TREP	21	16	12	30
OS	19	19	9	20
MKWS			19	

Table 4.
Frequency of Strategies in the Argumentative Genre

Strategy	Participant 1	Participant 2	Participant 3	Participant 4
TT	1		2	2
SR	23	23	45	27
ST	12			26
SDQ	10			10
OR	26	23	24	23
OT	10		1	17
TREP	7	22	16	27
OS	23	23	23	23
MKWS				

For the expository text, there were 19 questions and for the argumentative text, there were 23 questions. To answer each question, the participant read the stem once and for some questions more than once. This is shown by the frequency of SR. Also, the participant attempted to select the correction option(s) for each question, as indicated by the frequency of OS. Regarding each participant, as Tables 3 and 4 show, Participant 1 used self-directed questions three times while answering expository questions and 10 times while answering argumentative questions. However, he used text reprocessing 21 times while answering expository questions and seven times while answering argumentative questions. Participant 2 used stem translation five times and self-directed questioning once. However, she did not use this strategy while answering argumentative questions. She used text-reprocessing 16 and 22 times for expository and argumentative questions, respectively. Participant 3 first read the stems of all the questions and then returned to the beginning of the test, read questions again one by one, and answered them. She used stem reading 32 and 45 times for expository and argumentative questions, respectively. She also marked the keywords in stems while reading the stems related to the expository text the first time, hence 19 MKWS. Participant 4 used stem translation three and 26 times for the expository and argumentative genre, respectively. She also used self-directed question three times for the expository questions and option translation 17 times for the argumentative questions. She also employed text-reprocessing 30 and 27 times for the expository and argumentative questions, in that order.

The tables demonstrate that the strategies TT, ST, and OT had a low frequency. MKWS was used only by one participant. SDQ was utilized by some participants. As to SR, there was a completely negative relationship between the participant's score and the use of this strategy, i.e. The participant with the highest score used this strategy less and the one with the lowest score used this strategy more. Such a relationship suggested that weaker participants needed to read the stem more in order to check each option against it, whereas stronger participants had mastery over the stem by just reading once or more in a few cases. OR, TREP, and OS had a high frequency. Each option was an IU and this required the examinee to check its truth considering the text. Therefore, the

examinee was required to return to the text, hence the high frequency of OR and TREP

Discussion

The fuzzy test developed in this study gives us a fuzzy score constituted of two scores. One score is a measure of deep understanding and the other score is the measure of surface understanding. To gain these scores, IUs were used as the basic unit of measurement. The use of IUs (i.e. the same propositions) as the units comprising a text has advantages. By calculating the percentage of participants referring to a particular IU, we can assess the difficulty of that part of text. This can help us adapt texts for students of different levels, i.e. make the texts easier or more difficult by making changes in the IUs. Also, the analysis of IUs can indicate the amount of overlap among participants. This may give us clues regarding misunderstanding among readers of a particular text, written or spoken. In other words, those units least referred to are, in fact, problematic ones. Therefore, these parts can be improved to make the text more understandable. Also, referring to different IUs by different readers of the same text may reveal why different interpretations of the same text occur. In fact, readers may focus on different IUs.

This study compared fuzzy scores and TOEFL scores as binary scores. About 70% of participants had fuzzy scores higher than their binary scores. The reason may be the fact that, in common binary tests, test-takers do not have the possibility of gaining points for their partial knowledge while fuzzy tests provide such possibility. In fact, choosing one option does not prevent the test-taker from gaining points for those parts of the answer he or she knows. As a result, this framework can give us a more accurate picture of the test-taker than common multiple-choice tests. This possibility of showing partial knowledge and gaining scores for it may account for not having a very strong correlation coefficient (.432). If there was a perfect correlation, it would mean there is no difference between the two tests. In fact, both the TOEFL and the fuzzy test assess the same construct and score similarly. However, the correlation coefficient gained and the fact that most participants had fuzzy scores better than the TOEFL scores indicate the possibility of showing partial knowledge and being scored for it on the fuzzy test.

In order to investigate the processes participants go through to answer multiple-choice questions based on fuzzy logic, four participants were asked to think aloud while answering the questions. Although each participant dealt with questions differently, the trend in the analysis of verbal reports was insightful. In fact, as Cohen (1998) stated, strategies test-takers use while taking a test can give us insights into test reliability and validity. In fact, "respondents may be using test-wiseness to circumvent the need to tap their actual knowledge or lack of it, consistent with Fransson's (1984) assertion that respondents may not proceed via the text but rather around it" (Cohen, 1998, p. 92). In this study, respondents read the text and reprocessed the text again and

again in an effort to answer fuzzy items. Except for Participant 3 who deleted a few options because of the word "only" in them, other participants did not use test-wiseness strategies which Cohen and Upton (2006) mention as i.e. "using the process of elimination (selecting an option even though it is not understood, out of a vague sense that the other options couldn't be correct), using clues in other items to answer an item under consideration, and selecting the option because it appears to have a word or phrase from the passage in it-possibly a key word" (p. 37) or any other strategies other than reading or test-taking strategies. In fact, the test engages the test-taker in a process of reading the text and questions, and using strategies to reach the correct answer. The test-takers used their general understanding to answer some questions and their local understanding when they reprocessed the text to find specific information. This happened when participants reprocessed the text to mark the correct answer, i.e. when they used text-reprocessing strategy. Questions asking the MI of the text and the writer's idea required participants' understanding across the text, and those asking true/false questions, a special person's idea, or a particular subject, required participants' local understanding, either at sentence or paragraph level. Therefore, the fuzzy questions developed in this study to which test-takers had to answer by choosing as many correct options as there are assess the reading comprehension of the test-takers as they are expected to do. There is a similar study, i.e. Cohen and Upton (2006), which investigated strategies used to answer single-selection multiple-choice formats which are basic comprehension and inferencing questions and the new selected-response multiple-selection multiple-choice reading-to-learn items. They mentioned "the most recently available new TOEFL reading task specifications (ETS, 2003)" (p. 6) which resulted in three types of reading tasks designed specifically to focus on reading for basic comprehension tasks, reading-to-learn tasks, and inferencing tasks which have elements of both of the other types. After analyzing the verbal reports of 32 participants, they found that reading-to-learn and inferencing items did not require or assess approaches to academic reading differing from those elicited by the basic comprehension questions.

Conclusion

In this study, a new version of a multiple-choice reading comprehension test was developed based on fuzzy logic. To this end, IUs were extracted and, accordingly, fuzzy questions were developed. Each question had five options and each option was an IU. The validity was considered in terms of concurrent validity with the reading section of a TOEFL test and the test-taking strategies participants used. The results indicated that most participants had better scores in fuzzy scores than TOEFL scores. Also, the analysis of verbal reports revealed that participants used test-taking strategies to answer reading comprehension questions, not test-wiseness.

The developed test adopted a new approach to assessment. This approach aims to measure the reading comprehension of examinees in a way that is both fairer and can be administered to a large number of participants. It is fairer as it

considers the partial knowledge of the examinees. In fact, the participant's understanding of each IU was considered and scored. It is in contrast to common multiple-choice tests on which examinees lose marks even though they know part of the answer.

The unit of measurement was IUs which are propositions or basic units of meaning forming a text. This helps us find the easy and difficult parts of a text by calculating the percentages of participants understanding a particular IU. In this way, we can adapt texts to render them suitable for the classroom context. In fact, by finding problematic parts (i.e. IUs) of a text, we can change them for the purposes and to suit the level of the class.

References

Al-Hammadi, A. S., & Milne, R. H. (2004). A neuro-fuzzy classification approach to the assessment of student performance. *Proceedings of the 2004 IEEE International Conference on Fuzzy Systems, Hungary,* 837-841. doi:10.1109/FUZZY.2004.1375511

Andoline, C. (1980). Syntactic maturity and vocabulary richness of learning disabled children at four age levels. *Journal of Learning Disabilities, 13*(7), 27-32.

Andrews, R. (2010). *Argumentation in higher education: Improving practice through theory and research*. Abingdon, UK: Routledge.

Baba, A. F., Bakanay, D., & Cin, F. M. (2012). A fuzzy system for evaluating students' project in engineering education. *Computer Applications in Engineering Education*, *20*(2), 287-294. doi: 10.1002/cae.20395

Baba, A. F., Kuscu, D., & Han, K. (2009). Developing a software for fuzzy group decision support system: A case study. *The Turkish Online Journal of Educational Technology, 8*(3), 22-29.

Bardovi-Harlig, K. (1992). A second look at T-unit analysis: Reconsidering the sentence. *TESOL Quarterly*, *26*, 390-395.

Bardovi-Harlig, K., & Bofman, T. (1988, March). *A second look at T-unit analysis*. Paper presented at the 22nd Annual TESOL convention, Chicago, IL.

Budd, R. (1988). Measuring proficiency in using English syntax. *System*, *16*(2), 171-185.

Chafe, W. L. (1985). Differences between speaking and writing. In D. R. Olson, N. Torrance, & A. Hildyard (Eds.), *Literacy, language and learning: The nature of consequences of reading and writing* (pp. 105-123). New York, NY: Cambridge University Press.

Cohen, A. D. (1998). Strategies and processes in test-taking and SLA. In L. F. Bachman & A. D. Cohen (Eds), *Interfaces between second language acquisition an language testing research* (pp. 90-111). England: Cambridge University Press.

Cohen, A. D., & Upton, T. A. (2006). *Strategies in responding to the new TOEFL reading task.* Princeton, NJ: ETS.

Cin, F. M., & Baba, A. F. (2008, May). *Assessment of English proficiency by fuzzy logic approach.* Paper presented at the 8TH IETC, Eskişehir, Turkey.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education*. London, England: RoutledgeFalmer.

Distefano, P., & Valencia, S. (1980). The effects of syntactic maturity on comprehension of graded reading passages. *The Journal of Educational Research*, *73*(5), 247-251.

Faghih, N., & Alamdar Youli, F. (2007). *Manteqe fazi dar azmoonhaye chand gozine-ee* [Fuzzy logic in multiple-choice tests]. Shiraz, Iran: Rakhshid.

Gee, J. P. (1996). *Social linguistics and literacies: Ideology in discourses*. London, England: Taylor & Francis.

Ghonsooly, B. (2013). *Boroonfekani andishe dar maharate khandan dar zabane dovvom* [Thinking-aloud in second language reading]. Mashhad, Iran: Ferdowsi University of Mashhad Publications.

Harris, J. (2000). *An introduction to fuzzy logic applications*. Dodrecht, the Netherlands: Kluwer Academic Publishers.

Hosseinchary, M., & Yosefy, F. (2008). Barresiye boloughe nahviye neveshtary dar daneshamoozane doreye ebtedaee [The investigation of written syntax development in primary school students]. *Nashriye Daneshkadeh Adabiyat va Oloume Ensaniye Daneshgahe Shahid Bahonare Kerman, 24,* 305-323.

Kirszner, L. G., & Mandell, S. R. (2012). *Patterns for college writing: A rhetorical reader and guide (*12th ed.*).* Boston, MA: Bedford/St. Martin's.

Mariconda, B. (2001). *Step-by-step strategies for teaching expository writing.* New York, NY: Scholastic.

McNeill, F. M., & Thro, E. (1994). *Fuzzy logic: A practical approach*. Chestnut Hill, MA: Academic Press.

Rashidi, N., & Shahballa, A. H. (2010, July). *The contribution of the idea unit analysis to the measurement of the reader's meaning and author's meaning coincidence.* Paper presented at the 3rd Annual Seminar on Current Issues and Problems in L2 Reading and Writing, Shiraz, Iran.

Richards, J. C. (1983). Listening comprehension: Approach, design, procedure. *TESOL Quarterly*, *17*, 219-240.

Roll, M, Frid, J., & Horne, M. (2007). Measuring syntactic complexity in spontaneous spoken Swedish. *Language and Speech*, *50*(2), 227-245.

Samim-Banihashemi, A. (1992). *Motale'ye roshde nahviye neveshtari dar miane daneshamoozan [The investigation of written syntax development among students]* (Unpublished MA thesis). Shiraz University, Shiraz, Iran.

Shahballa, A. H., & Alamdar Youli, F. (2012). Reading comprehension of different genres: A fuzzy approach. *International Journal of English Linguistics, 2*(1), 17-27.

Zadeh, L. A. (2008). Is there a need for fuzzy logic? *Information Sciences, 178*, 2751–2779.

Zahedi, H. (2001). *A fuzzy logic approach toward appreciating partial knowledge in multiple-choice testing* (Unpublished master's thesis). Shiraz University, Shiraz, Iran.