

The Study of Application of Simple Clitics in Sports Media Headlines

Marziyeh Dehghan Niri¹; Faezeh Arkan² 

1. PhD in Linguistics, English Language and Literature Department, Humanities faculty, Islamic Azad University, Qom branch, Qom, Iran.

marziyeh_dn@yahoo.com

2. Assistant Professor, Foreign Languages Department, Humanities Faculty, Hazrat-e Masoumeh University, Qom, Iran.

farkan@hmu.ac.ir (corresponding author)

Abstract

This study examines three functional morphemes which can be realized as clitics, in the sports newspapers headlines. Clitics are bound morphemes which are divided into simple and special clitics each of which has several characteristics. The aim of this study is to examine simple Persian clitics and their frequency in the sports newspapers headlines. To do so, the headlines of seven sports newspapers (namely, Abrar-e Varzeshi, Piruzi, Esteghlal-e Javan, Khabar-e Varzeshi, Goal newspaper, Navad 90, and Iran-e Varzeshi) were collected over a two-month period (May and June 2014) and their simple clitics were analyzed based on Shaghghi's (1995) classification. The results showed, in spite of our expectation, that simple clitics were rarely used in the prose of sports newspapers headlines and the frequency of their full form, that is, functional morphemes of "ra, va, and ham", were significantly higher than the frequency of their corresponding clitic forms. This was probably due to the higher fluency and ease of comprehension of the headlines by the readers of the sports newspapers.

Keywords: Clitic, Simple clitic, Enclitic, Language and sports media

Citation: Dehghan Niri, M., & Arkan, F. (2024). *Applications of Language Studies*, 2(2), 1-19.

1. Introduction

A clitic is a grammatically unaccented bound morpheme that can be used in structures larger than words. It forms a phonological word together with its base lexeme. Phonological words are syntactic groups or structures that are phonologically pronounced like words but are not considered as words in terms of morphology, syntax, and semantics (Shaghaghi, 2009, p. 74). Clitics, like affixes, need a host to be attached to. However, unlike affixes, they are not considered as part of a word structure (Shaghaghi, 2009, p. 73). Clitics can be divided into four categories depending on their position in relation to the host word they are connected to: “proclitic” and “enclitic”, which are attached to the beginning or end of the host word, respectively. Furthermore, according to Harris (2002) and Kari (2003), there are “mesoclitics”, which appear between the host word and other affixes, and “endoclitics”, which split the base of the host word in two parts and are inserted between them (Jam, 2013, p. 62).

According to Castairs-McCarthy (1992, p. 142), clitics are syntactically similar to words and phonologically and morphologically similar to affixes. Zwicky (1994, p. 572) says that a clitic is an element that has similar properties to words but is not independent like them; it does not appear in a sentence by itself and is phonologically dependent on the neighboring word.

Shaghaghi (2012), in her “phrasal affixes”, analyzes the position of clitics in the Persian language structure. Accordingly, she considers clitics to belong to phrasal affixes or to the second-position ones. Those located at the edge of the group are phrasal affixes, while those that are not dependent on category of the base are second-position clitics.

Clitics are among the elements found in many Iranian languages and show interesting phonological, morphological, and syntactic features that need to be studied. Although much research has been conducted on these clitics, a deeper understanding of their form and function can be gained by looking at them from a cognitive perspective. The discussion of clitics has emerged in linguistic debates since 1970s. Clitics are placed somewhere between affixes and words. Some criteria have been proposed to distinguish them from the other two groups. Clitics are usually unaccented and depend on the host word to which they are

attached; they appear on the final level of the word and no other affixes may be added afterwards; they are free to choose their host; they may go under syntactic movement and they are rule-governed. Clitics are divided into simple and special types. Simple clitics are reduced phonological forms of independent morphemes and usually have the same distribution as their free forms. Special clitics do not necessarily have the same phonological forms as their counterparts and there are certain rules for their distribution. One of the most important classes of clitics is the pronominal clitics which are found in various languages including the Iranian languages, and have attracted great attention of linguists due to their interesting features (Rasekh Mahand, 2013).

Simple clitics in Persian are reduced forms of the functional morphemes 'ra', 'va', and 'ham', which are commonly used in fast speech forms and casual conversations (Shaghghi, 2014). This study investigates the frequency of simple clitics in 343 print sports media headlines. Though the language of sports newspapers is much closer to spoken language and colloquial speech than other newspapers, we should examine the rate of frequency of simple clitics in sports newspaper headlines and also their similarities and differences with spoken language in using simple clitics. For this purpose, within the framework of Shaghghi's classification (1995), this type of clitics is studied and analyzed. Specifically, the questions of this study are as follows:

1. How frequent are simple clitics in sports newspaper headlines?
2. What are the similarities and differences between using simple clitics in sports newspaper headlines and spoken language?

2. Review of Literature

This section describes the most relevant previous research on clitics. It begins with studies conducted outside Iran on clitics including Jacob Wackernagel (1892), Spencer (1991), Zwicky (1977, 1985, 1994), Klavans (1985, 1982), Katamba (1993), Katamba and Stonham (2006), Gerlach (2002), Kruger (2005), Booij (2005), Katamba and Stoneham (2006), Gerlach (2002), Kroeger (2005), and Booij (2005) and continues with studies carried out in Iran including Shaghghi (1995, 2007, 2010, 2013), Meshkat-o-Dini (2000), Kalbasi (2001),

Rasekh Mahand (2005, 2010), and other Iranian studies. In Indo-European languages, clitics appear in second position after an accented element in a sentence (Wackernagel, 1892).

Clitics are elements with some independent properties, but they are usually distinct from independent words; i.e. they cannot stand alone. They are phonologically attached to a host word. Clitics are usually grammatical words such as pronouns or auxiliary verbs. Historically, they are often derived from full words and eventually become inflected affixes. The set of phenomena known as cliticization is a meeting point of morphology, syntax, and phonology (Spencer, 1991).

Historically, the label “clitics” was originally applied by ancient Greek grammarians to elements that were syntactically words but were morphologically related to the neighboring words (specially the preceding words). Together with their base words, they form a kind of morphological unit (Zwicky, 1994). In describing clitics, Zwicky (1977) argues that in most languages there are morphemes that present analytical difficulties because they can be clearly classified neither as independent words nor as affixes (Zwicky 1977). He believes that taxonomic and theoretical assumptions about clitics require distinguishing them from words and inflectional affixes. It can therefore be found that the issue of distinguishing between words, affixes, and clitics is of great importance. A clitic is an element that has some of the properties of a full word but, unlike a word, lacks independence (Spencer, 1999). For this reason, Zwicky made great efforts to classify clitics. In this context, Zwicky (1977, 1985), Klavans (1982, 1985) and Zwicky and Pullum (1983) presented a set of tests and parameters that can be used to identify and classify clitics. Based on Zwicky and Pullum's tests (1983), bound morphemes can be divided into four groups: derivational affixes, inflectional affixes, simple clitics, and special clitics. The connection of a clitic with its base and its realm of function can be determined using the parameters proposed by Klavans. The first parameter determines the base or host of the clitic so that the clitic attaches to the first or last syntactic daughter constituent in the sentence and is called the initial or final constituent, respectively. The second parameter depends on the position of the clitic in relation to this syntactic daughter which may appear before or after that. The third parameter concerns the

direction of the clitic attachment which is connected to the base as a proclitic or enclitic (Shaghghi, 2012, cited in Klavans).

Katamba (1993) defines clitics as bound morphemes that can be attached to independent words by syntactically motivated rules. He calls these independent words as hosts. Clitics may appear at the beginning or at the end of their hosts, which are called proclitics or enclitics (Katamba, 1993).

Katamba and Stonham (2006), citing Zwicky (1997), view clitics as bound morphemes which occur almost in all world languages. Neither are they independent words, nor do they belong to affixes. They also state that, in addition to inflectional affixes, there is another class of bound morphemes which are called clitics. Clitics can attach to independent words using syntactically motivated rules. The word to which clitics are attached is called the host.

Some recent research suggests that clitics should be classified as a 'word' or 'affix'. Others propose the category of “phrasal affixes” to refer to clitics (Gerlach, 2002). As defined by Krueger (2005), clitics are linguistic forms that resemble both independent words and affixes. Booij (2005, pp. 168-166) considers clitics as “reduced words that refer to function words or non-lexical classes which depend on the host word used before or after them, such as pronouns and determiners, and they cannot be considered as phonological words by themselves” (Sarahi & Alinejad, 2013, p. 106).

According to Iranian traditional grammarians, when linguistic elements such as “-am, -at, -ash, -eman, -etan, -eshan” function as an object, they are called bound object pronouns (Qarib et al., 1987, pp. 74-77; Natel Khanlari, 1980, pp. 198-199; Ahmadi Givi & Anvri, 1995, pp. 30-31). Later studies, however, have a different view on these elements and call them as clitics. Meshkat al-Dini (2000) believes that although they refer to a direct object (in terms of grammatical relations), they do not bear primary accent and are not part of the inflectional morphology; therefore, they are clitics. In Kalbasi's (2008, p. 25) view, a clitic is a kind of word that cannot stand alone and, like affixes, is attached to the preceding or following word; but, unlike affixes, it is not considered part of the structure of that word.

Clitics that are attached to the end of a word are “enclitic” and clitics that are attached (added) to the beginning of a word are “proclitic”. Persian clitics are mostly enclitic.

Clitics are unaccented grammatical morphemes that do not appear in structures larger than words and phonologically form phonological words with their base. “A phonological word is a syntactic group or constituent that is phonologically pronounced like a word, but is not a word in terms of morphology, syntax, and semantics” (Shaghaghi, 2016, p. 74). In defining clitics, Shaghaghi (2010, p. 73) also writes, “in their researches in the second half of the 20th century, grammarians encountered dependent morphemes that played a grammatical role but were not included in the category of affixes. These morphemes were not involved in the production of words or word forms, but they participated in the field of syntax, i.e., in the construction of syntactic groups and sentences”. She adds, “on the one hand, clitics follow syntactic rules as free lexical and grammatical morphemes and words, but on the other hand, they behave like affixes because of their dependent nature; i.e. they necessarily attach to their bases. Unlike affixes, these elements are not considered part of the word structure and cannot be dropped”.

Analyzing the position of clitics in Persian syntax, Shaghaghi considers that clitics may belong to phrasal affixes or to the second-position. Those that attach at the margins of the group are group affixes, while those that appear free from base are second-position clitics (Shaghaghi, 2013).

Descriptive studies of clitics show that this element resembles words in some features and affixes in others. Zwicky (1994) notes that the categories of words and affixes are distinct in terms of syntax, morphology, semantics and phonology. Words have a special syntactic category and can potentially be the head of a group to which modifiers are added. Words are lexical morphemes and have independent stress; while affixes do not have a syntactic category and cannot be the head of a group; their meaning is based on syntactic relations and they attach to a morphological base. In the sentence /Gav ælæfha ra khord/ (The cow ate grass), the four morphemes "gav", "ælæf", "ra" and "khord" are examples of words, while the grammatical plural morpheme "-ha" is an example of an affix. Clitics, by their nature, have some of the features of words and some of the features of affixes. There is actually a

continuum between words at one end and affixes at the other end. Clitics are divided into three categories: simple clitics, special clitics, and bound words. Reduced pronouns in English are simple clitics (e.g., in the sentence “give-m to me”, “-m” is the reduced form of “them”). Simple clitics share the same distribution with their full forms and they are clearly related to their full forms semantically and phonetically. For example, in Persian, the enclitic /-o/ is the reduced form of “ra”. The morpheme "ra" is pronounced as "/ro/" in colloquial speech and is reduced to "/-o/" when preceded by a word which ends in a consonant letter (Shaghghi, 1995).

Klavans (1985) argues that it is possible to provide a single theory of clitics, but special distinctions must be taken into account in this regard. These distinctions are in the form of parameters that can indicate similarities and differences between clitic systems in different languages. On the other hand, she also emphasizes that different types of clitics in a given language may indicate different parameters. Klavans claims that what distinguishes clitics from words or affixes is that, unlike affixes, clitics (usually) attach syntactically to a particular group, and this group is considered as the realm of clitics function. Therefore, Klavans uses the term “phrasal affix” to refer to clitics in this context (Spencer, 1991, p. 377). Considering the parameters proposed by Klavans, it can be said that they all appear after the base and are enclitic according to the direction of connection. Each clitic is dominated by one or more groups.

Within the framework of Shaghghi's classification (1995), which is the theoretical background of this study, she (1995) identifies three simple and six special clitics out of over a hundred bound morphemes analyzed. In all the nine cases mentioned above, clitics are unaccented and bound, which prevents the application of morphological affixation rule and there is no accidental gap in any of the cases. The distribution is similar to that of words and the ellipsis rule applies to all clitics except for the ezafe kasre and indefinite article '-i'; while the movement rule cannot be applied to any clitic. In other words, according to the accidental gap test, the distribution and ellipsis rules, clitics are similar to words; whereas considering the dependency test, affixation avoidance and movement rule, they resemble affixes, especially inflectional affixes. The dependency test also shows no similarity between clitics

and words and (derivational or inflectional) affixes; thus, they exhibit an intermediate status between words and affixes.

Simple clitics in Persian are reduced form of the words 'ra', 'va' and 'ham', which are commonly used in fast speech forms and casual conversations. Special clitics can be divided into two categories: clitics that do not have a full form such as indefinite article '-i', indefinite determiner '-i', ezafe kasre, and emphasis marker '-ha'; and clitics that have a full form but do not have the same distribution such as the bound stems of the verb 'to be' and bound personal pronouns (Shaghaghi, 1995, p. 150). Therefore, the study and analysis of the use of simple clitics in the titles of sports print media are conducted based on Shaghaghi's classification (2014). In this respect, the use of simple clitics is assumed and expected to be relatively higher than the use of them in other print media because the language variety used by this type of media is closer to the language commonly used in colloquial speech. It seems that clitics occur more frequently in this language variety, but it remains to be seen what the results of the study will be.

3. Method

In this study, we examine three functional morphemes which can be realized as simple clitics in print sports media headlines. This study is almost considered as a quantitative descriptive research study. In general, the data in this study was collected from a small corpus, that is, sports newspapers including Abrar-e Varzeshi, Piruzi, Esteghlal-e Javan, Khabar-e Varzeshi, Goal newspaper, Navad 90, and Iran-e Varzeshi. They were selected from May and June 2014; each newspaper office usually published 25 newspapers in May 2014 and 24 newspapers in June 2014. The method was based on selecting headlines of these newspapers, counting the number of simple clitics, and then comparing them.

In the analysis of every headline, functional morphemes including ra, va, and ham were observed and then calculated together with their clitic forms, which can be realized as simple clitics. In sum, a total of 343 headlines (175 in May and 168 in June) were selected and these grammatical morphemes, as the data for this study, were checked and counted in these headlines, then their frequency was considered comparing to each other. More

importantly, the aim of this study was to examine if the simple Persian clitics are used in the headlines of sports newspapers as the print media and how much their frequency was.

4. Results and Discussion

In this section, we will examine simple clitics in print sports media headlines in the above newspapers. In practice, these data consists of simple clitics that are considered as bound morphemes, which have the same syntactic distribution as their full forms as well as clear semantic and phonetic associations with them (Shaghaghi, 1997).

The number of grammatical morphemes that can be used as simple clitics in the spoken language is calculated in the headlines and shown in the following tables.

Table 1

The Number of Grammatical Morphemes that Can be Used as Simple Clitics in Newspaper Headlines in May

Newspaper	/ra/	/væ/	/hæm/
(May)			
Abrar-e Varzeshi	1	3	1
Piruzi	5	1	1
Esteghlal-e Javan	3	1	3
Khabar-e Varzeshi	5	4	-
Goal newspaper	1	2	-
Navad 90	3	-	-
Iran-e Varzeshi	-	1	-
Total	18	12	5

Table 2

The Number of Grammatical Morphemes that Can be Used as Simple Clitics in Newspaper Headlines in June

Newspaper (June)	/ra/	/væ/	/hæm/
Abrar-e Varzeshi	-	3	2
Piruzi	5	1	1
Esteghlal-e Javan	5	2	2
Khabar-e Varzeshi	4	2	-
Goal newspaper	2	3	-
Navad 90	4	2	-
Iran-e Varzeshi	-	-	\
Total	20	13	6

Table 3

The Number of Grammatical Morphemes that Can be Used as Simple Clitics in Newspaper Headlines in May and June

Newspaper	/ra/	/væ/	/hæm/
May	18	12	5
June	20	13	6
Total	38	25	11

Table 4

The Total Number of Headlines and Total Number of /ra/, /væ/, and /hæm/

Newspaper	Total number of headlines	Total number of /ra/, /væ/, /hæm/
May and June	343	74

Table 4 shows that there are only 74 words in about 343 headlines in the two-month period that can be used as clitics, but only in two headlines among all these cases, the simple clitic (bound) forms were used instead of its full form:

1- /Parvin: in pesare dige 'kard \mathbf{o} ' (ka:rd- \mathbf{o}) be ostokhun resunde! (Goal newspaper, Wednesday, April 29, 2015). (Parvin: I am fed up with this guy!)

This is the first headline in which the clitic "o" instead of "ra" is used in the second position, after the noun in this noun phrase, and it is added to the base noun 'ka:rd- \mathbf{o} ' (knife), which is used idiomatically rather than denoting a definite noun. Moreover, according to Halpern (1998), the string 'ka:rd \mathbf{o} ' is not necessarily a semantic structure but a phonological word that appears in the second position by attaching the clitic \mathbf{o} (o) to the base noun. The second headline in which the clitic $\mathbf{-o}$ instead of "ra" is used is as follows:

2- /Amrica ra bezanid **va halesho** bebarid/ (Esteghlal, Thursday, June 18, 2015)
(Defeat US team and live it up.)

What is important in the headline "Amrika **ra** bezanid va halesho bebarid" is the use of two clitics "-esh" instead of "a:n" (that) and "o" instead of "ra" : (halesho = hal-e a:n **ra**). It should be added that in the pronunciation of hal-e sho, the clitic $\mathbf{-e}$ is also uttered, but in the prose/ writing, this $\mathbf{-e}$ is not written. Only $\mathbf{-o}$ is a simple clitic and $\mathbf{-sh}$ and $\mathbf{-e}$ are not simple clitics, but they are pronominal and special clitics, respectively. In another headline, "Perspolis!omideman **ra** naomid nakon": (omideman =omid-e ma), "-eman/ -ema:n" here is a pronominal clitic which refers to "ma", while "**ra**" is independent. In this case, "-eman" as a clitic is attached to the base, though it is not considered as a simple clitic and it is not the subject of this study. Generally, these clitics "-eman", "-e", "-esh", and "-o" are enclitics which appear at the end of nouns, although they do not belong to the same category of clitics.

In other headlines (342 out of 343 ones), the words 'ra', 'va', and 'ham', which can be represented as clitics, are used in their full forms in the prose of these sports newspapers, a few examples of which are mentioned here:

3- /perspolis ba Branko **ham** nemi-barad/ (Abrar-e Varzeshi, Monday, April 27, 2015)

(Persepolis can't win with Branko, either.)

4- /Perspolis! Omideman **ra** naomid nakon/ (Piruzi, Sunday, April 26, 2015)

(Persepolis! Don't let us down)

5- /ham teractor ra mibarim **ham** Perspolis ra/ (Esteghlal-e Javan, Saturday, May 9, 2015)

(We'll win both Tractor and Persepolis)

6- /Ladrop moch-e Branko **ra** khaband/ (Khabar-e Varzeshi, Thursday, April 23, 2015)

(Laudrup beat Branko)

7- /Branko: Daei **va** Karimi rofaghayam hastand. Dust daram biyayand sar-e tamrin/
(Goal newspaper, Wednesday, May 6, 2015)

(Branko: Daei and Karimi are my friends; I want them to attend the training sessions)

8- /nemitavanam ba pul-e khodam esteghlal **ra** bekharam/ (Navad 90, Tuesday, April 28, 2015)

(I cannot buy Esteghlal with my own money)

/Bazi-ye marg **va** zendegi/ (Iran-e Varzeshi, Wednesday, May 6, 2015)

(A game of life and death)

9- /esteghlal emruz **ham** malek peyda nemikonad/ (Abrar-e Varzeshi, Monday, May 25, 2015)

(Esteghlal can't find an owner today, either)

10- /Branko: bekhahand tim **ra** baraye qahremani mibandam/ (Piruzi, Thursday, May 28, 2015)

(Branko: If they want, I will arrange the team for championship)

11- /ta esteghlal hast, Hejazi **ham** khahad bud/ (Esteghlal, Saturday, May 23, 2015)

(Hijazi will exist as long as Esteghlal exists)

12- /Karimi: in aqayan **va** in football/ (Khabar-e Varzeshi, Tuesday, June 2, 2015)

(Karimi: these men and this soccer)

13- /Daei va Karimi **ra** ashti midahim/ (Goal newspaper, Tuesday, June 16, 2015)

(We'll help Daei and Karimi be reconciled)

14- /shartbandi-ye yeki az nazdikan-e Daei Qahramani **ra** gereftar kard/ (Navad 90, Monday, June 1, 2015)

(A close relative of Daei put Qahramani into trouble by betting!)

15 /porfosor **ham** ruye hava/ (Iran-e Varzeshi, Monday, June 1, 2015)

(Professor up in the air, too!)

5. Conclusion

Simple clitics in Persian are the reduced forms of the words 'ra', 'va', and 'ham' and are usually used in fast and casual speech (Shaghghi, 1995). Thus, we studied the use of simple clitics in the headlines of print sports media because the languages used in this type of media is closer to colloquial speech than other media. Therefore, it is expected that simple clitics are used more frequently in this type of media than others. In spite of this expectation, our study showed that in all cases where simple clitics might be used, their full forms were used except for two cases (ka:rd-**o** and hal-e-sh-**o**), the dependent form "-o" is used instead of "ra". Totally, the data in the above sentences and their statistics in the tables show that among the words 'ra', 'va', and 'ham', 'ra' is the most frequent functional word, which was used 38 times; thus, it has been assumed that its clitic form might be more frequently used than 'and' with the occurrence of 26 times and 'ham' with the occurrence of 11 times. Only two occurrences of the simple clitic "-o" instead of "ra" were observed in the data. Thus, in answering the first question, it was found that the frequency of simple clitics in print sports media is very low and can be ignored. In other cases, even if the full form is used in writing, it is pronounced as

its dependent form in reading. In most cases, this applies to the word "va" which is pronounced as "-o".

Simple clitics are basically used in spoken language not in written language, but sometimes they are used with very low frequency in sports newspapers and the prose of these sports media tends to be similar to colloquial speech. Furthermore, since the readership of sports newspapers comes from different walks of life (ranging from children, adolescents to adults), on the one hand, it seems that a fluent writing style which is close to the colloquial speech is required at first glance. On the other hand, the lack of higher number of simple clitics in sports newspaper headlines shows the fact that sports newspaper writing or prose is not very close to colloquial speech and has its own writing rules. Therefore, it can be concluded that, for fluency and correct understanding, the full forms of functional words "ra", "va" and, "ham" are used even in less formal sports newspapers comparing with other more formal newspapers like political ones, although the readers of which may be of a rather higher level of formality.

The results showed that, in spite of our expectation, due to the more fluency and proper comprehension of the text, simple clitics are rarely used in the prose of print sports media headlines and the frequency of the full forms of the functional morphemes of 'ra, va, and ham' is significantly higher than the frequency of their corresponding clitic forms.

From the theoretical point of view, simple clitics are fundamentally belonging to the spoken language, not the written one. In line with this linguistic fact, the findings show that headline writers of Persian sports media follow and observe this fact. Thus, they do not have a tendency to use simple clitics in their headlines, although utilizing this type of clitic is expected because the prose of these media is similar to the spoken language. Summing up, in answering the second question, it should be mentioned that there is a significant difference between using simple clitics in the print sports media and the spoken language.

The very limited use of simple clitics as the bound morphemes in headlines paves the way for the next researches to consider other types of grammatical morphemes in the newspapers in order to study their frequency and application in the prose of sports newspapers. In addition, the comparative studies on how to use clitics and other grammatical

morphemes such as prepositions in sports headlines could be a research topic for future researchers.

Conflict of interest

The author(s) certify/certifies that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in the present research paper.

References

- Ahmadi Givi, H., & Anvari, H. (1995). *Persian Grammar1*. Tehran: Fatemi. Doi: <https://doi.org/10.30479/JTPSOL.2023.17466.1595>. [in Persian]
- Bahrami, F. (2012). *Object indexers: Clitics or agreement markers?* Proceedings of the First National Conference on Clitics in Iranian Languages, Compiled by Muhammad Rasekh Mahand, Tehran: Parsi Nevisesh Publishing House, pp. 79-98. [in Persian]
- Booij, G. (2005). *The Grammar of Words*, Oxford: Oxford University Press.
- Carstairs-McCarthy, A. (1992). *Current Morphology*, London: Routledge. London and New York: Routledge.
- Dabir Moghadam, M. (1990). About "ra" in Persian language. *Journal of Linguistics*, 5(1), 2-60. Doi: <https://doi.org/20.1001.1.20086261.1389.2.2.1.5>. [in Persian]
- Gerlach, B. (2002). *Clitics between Syntax and Lexicon*. Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- Halpern, A.L. (1998). *Clitics*. in A. Spencer & A.M. Zwicky(eds.), *Handbook of Morphology*. London: Blackwell.
- Harris, A. C. (2002). *Endoclitics and the Origins of Udi Morphosyntax*. Oxford: Oxford University Press. [in Persian]
- Jam, B. (2013). On Ezafe ([e]) Enclitic and Stress in Pronouncing Persian Last Names. *Journal of Adab Pazhuhi*, 30, 61-80. Doi: <https://doi.org/20.1001.1.17358027.1393.8.30.3.7>. [in Persian]
- Kalbasi, I. (2001). *Construction of Derivative Words in Modern Persian*. Tehran: Research Institute of Humanities and Cultural Studies. [in Persian]
- Kari, E. E. (2003). *Clitics in Degema: A Meeting Point of Phonology, Morphology, and Syntax*. Tokyo: Research Institute for Languages and Cultures of Asia and Africa.
- Katamba, F. (1993). *Morphology*. Macmillan Press LTD.

- Katamba, F. J. S. (2006). *Morphology*, 2nd ed., Houndmills/ London: Macmillan.
- Klavans, J. L. (1982). *Some problems in a theory of clitics*. Bloomington: Indiana University Linguistics Club.
- Klavans, J. L. (1985). The Independence of Syntax and Phonology in Cliticization. *Language*, 61, 95-120. Doi: <https://doi.org/10.2307/413422>.
- Kroeger, P.R. (2005). *Analyzing Grammar: An Introduction*, Cambridge University Press.
- Marantz, A. (1988). Clitics, morphological merger, and the mapping to phonological structure, in Michael Hammond and Michael Noonan (eds.), *Theoretical morphology: approaches in modern linguistics*. San Diego: Academic Press, pp. 253-70.
- Mazinani, A., & Sharifi, Sh. (2015). Investigation of the pronominal clitic system in the course of Persian history and the causes of its change. *Linguistic Studies*, 6(4), 275-305. [in Persian]
- Meshkat-o-Dini, M. (2000). *Persian Grammar Based on Theory of Transformation*. 2nd edition, Mashhad: Ferdowsi University. [in Persian]
- Nateli Khanlari, P. (1980). *Persian Grammar*. 4th and 5th editions, Tehran: Tus. [in Persian]
- Poorebrahim, Sh. (2004). *Comparative study of pronominal clitics in Delwari and Dashti dialects*. Proceedings of the first national conference on the study of clitics in Iranian languages. Compiled by Muhammad Rasekh Mahand, Tehran: Parsi Nevisesh Publishing House (2009). pp. 221-334. [in Persian]
- Qarib, A.A., Bahar, M.Al., Foruzanfar, B., Homaei, J., & Yasemi, R. (1987). *Persian Grammar (Panj Ostad)*. Tehran: Farhang. [in Persian]
- Rasekh-Mahand, M. (2005). *Object agreement in Persian*. Proceedings of the first conference of the Iranian Linguistics Association. Compiled by Mustafa Aasi, pp. 275-285. [in Persian]

- Rasekh-Mahand, M. (2007). *Description of Enclitic compound verbs and their entries in Persian Dictionaries*, Dictionary Writing, special edition of Farhangistan, vol. 1, pp. 253-236. [in Persian]
- Rasekh-Mahand, M. (2010). *Clitics as a sign of subject agreement in Persian*. Proceedings of the first national conference on the study of clitics in Iranian languages. Compiled by Muhammad Rasakh Mahand, Tehran: Parsi Nevisesh Publishing House, pp. 61-78. [in Persian]
- Rasekh-Mahand, M. (2010). Persian Clitics Adjacent to the Verbs, *Journal of Linguistics Studies*, 2, 75-85. Doi: <https://doi.org/20.1001.1.20086261.1389.2.3.5.1>. [in Persian]
- Rasekh-Mahand, M. (2012). *Examining the Clitics in Iranian Languages*. Tehran: Parsi Nevisesh Publishing House. [in Persian]
- Sarahi, M. A., & Ali Nejad, B. (2013). Classification of clitics in Persian. *Journal of Linguistics and Dialects of Khorasan, Ferdowsi University of Mashhad*,1, 103-130. Doi: <https://doi.org/10.22067/lj.v5i8.34489>. [in Persian]
- Shaghaghi, V. (1995). *What is Clitic? Is such a concept used in Persian language?* Proceedings of the third conference of linguistics, (February 5-6, 1996), Compiled by Yahya Modaresi and Muhammad Dabir Moghadam, Research Institute of Humanities and Cultural Studies and Allameh Tabatabai University, pp. 141-157. [in Persian]
- Shaghaghi, V. (1997). *The second position clitic in Persian*. Proceedings of the first national conference on the study of clitics in Iranian languages. Compiled by Muhammad Rasekh Mahand, Tehran: Parsi Nevisesh Publishing House, (2010), pp. 13-35. [in Persian]
- Shaghaghi, V. (2010). *An Introduction to Morphology*. Tehran: The Organization for Researching and Composing University textbooks in the Humanities (SAMT). [in Persian]
- Shaghaghi, V. (2013). Phrasal affix. *Journal of Language and Linguistics*, 9 (17), pp. 1-26. [in Persian]

Spencer, A. (1991). *Morphological Theory*. Basil Blackwell, Cambridge.

Tajvidi, Gh. R. (2004). *English Morphology Plus Word-Formation Processes in English & Persian*. SAMT Publications. Center for Humanities Research and Development. [in Persian]

Wackernagel, J. (1892). Uber ein Gesetz der indo-germanischen wort-stellung. *Indogermanische Forschungen*, 1, 333-436. Doi: <https://doi.org/10.1515/9783110242430.333>.

Zwicky, A. (1977). *On Clitics*. Bloomington: Indiana University Linguistics Club.

Zwicky, A. (1994). Clitics. *Encyclopedia of Language and Linguistics*. Edited by R.E, Asher Pergamon Press, vol.2, pp. 571-576.

Zwicky, A. M., & Pullum, G. K. (1983). Cliticization vs. inflection: English n't. *Language*, 502-513. Doi: <https://doi.org/10.2307/413900>.

Item Response Theory and Mantel-Haenszel Procedures in Detecting Academic Discipline Differential Item Functioning and Differential Skill Functioning with English Proficiency Test

Hamed Ghaemi¹  Mahsa Khorami²

1. Assistant Professor, Bahar Institute of Higher Education, Mashhad, Iran.
ghaemiacademy@gmail.com (corresponding author)

2. Instructor, Bahar Institute of Higher Education, Mashhad, Iran.
mahsa_m6669@yahoo.com

Abstract

The present mixed-method study aimed at investigating the presence of Differential Item Functioning (DIF) and Differential Skill Functioning (DSF) in a high-stakes language proficiency test in Iran, the English Proficiency Test (EPT) with different academic backgrounds (i.e., Humanities vs. Sciences & Engineering) using Item Response Theory (IRT) and Mantel-Haenszel (MH) approaches. It also aimed at detecting if there is any correlation between IRT and MH methods and also detecting which DIF items are biased. The English subtest consisted of a total of 100 items. The participants (N = 642) were selected by convenience sampling from universities in Tehran. The results displayed DIF between Sciences and Humanities students, but they did not show DSF in favor of a particular academic discipline group. Hence, on the basis of the findings, it was concluded that the EPT scores are not free of construct-irrelevant variance but because 14 DIF detected items out of 100 items were too small, the overall fairness of the test was confirmed. In addition, it was found that some of DIF detected items were bias and some of them functioned differently, simply because the two groups differed in their abilities. The positive correlation between IRT and MH methods was also proved.

Keywords: Differential Item Functioning, Differential Skill Functioning, Item Response Theory, Mantel-Haenszel, English Proficiency Test

Citation: Ghaemi, H., & Khorami, M. (2024). *Applications of Language Studies*, 2(2), 20-58.

1. Introduction

Reliability and validity are crucial goals in test formulation and design. The degree to which a test measures what it is supposed to assess is referred to as test validity. Validity is a crucial concept because without it, any measurement, inference, or conclusion based on it is useless. It is a truism, according to Geranpayeh and Kunnan (2007), that tests must be fair to test takers in order for test-score interpretations made by test users (e.g., admission officers and employers) to be valid. Quality management in test design, administration, and scoring, adequate coverage of relevant content, sufficient construct validation work, equal learning opportunities and access to testing, and items measuring only the skill or ability under investigation without being unduly influenced by construct-irrelevant variance introduced through test-taker bias, are all examples of test fairness (Saville, 2005; Shohamy, 2000).

When standardized English-language examinations are given to people all over the world, the test-taking population may differ based on personal and educational criteria including age, gender, first language, and academic discipline. Test developers must constantly check their tests to ensure that all test takers are receiving a fair test (Geranpayeh & Kunnan, 2007). Items must be appropriate for readers from a variety of racial, ethnic, cultural, and linguistic backgrounds in order for the exam to be valid (Abedi et al., 2005). So, test validation through various ways is a vital step for fair assessment of examinees as well as proper interpretation of a test, especially when considering the personal and social repercussions of a test.

The process of validating a test begins at the development stage and continues after it has been used. The validation process, according to Zumbo (1999), is never complete. As Bachman (2004) writes:

Language examinations have become ingrained in both our educational system and our culture. Language test results are used to make assumptions about people's language abilities and to inform decisions we make about them. Language exams, for example, are used to identify second and foreign language learners in schools, to choose students for university entry, to place them in language programs, to screen possible immigrants and to select

employees. As a result, language exams have the potential to assist us in gathering relevant data that will benefit a wide range of people. However, in order to actualize this promise, we must be able to demonstrate that the scores we acquire from language exams are accurate, as well as the methods we employ to evaluate and use those scores. (p. 3).

Differential Item Functioning (DIF) analysis, according to Li and Suen (2012), is one technique to analyze the appropriateness and validity of tests. It was originally developed to uncover possible item bias across subgroups. “A disparity in item performance between two comparable groups of examinees” is what DIF stands for (Dorans & Holland, 1993). It exists when the chance of successfully answering a question varies between groups (Koo et al., 2013). De Ayala et al. (1999) and Klieme and Baumert (2001) have suggested that multidimensionality may be the origin of DIF, and that DIF happens when one of the unexpected additional dimensions reflected by a score is connected to race/ethnicity, gender, or other demographic characteristics (Li & Suen, 2012).

According to Penfield (2001), the inclusion of DIF items may imply a systematic invalidity of the items, putting one group at a disadvantage. If there is a score discrepancy across groups as a result of unequal item performance, the validity of the scores may be jeopardized (Holland & Wainer, 1993). According to Park and French (2013), the existence of DIF items in an instrument indicates that the score measured by the instrument should be understood and used with caution because it could have various meanings for different group members. As a result, detecting, revising, or removing DIF items from highly selective tests is critical (Yu et al., 2006).

In the context of DIF, many studies have been conducted to identify group differences in performance and to see if test items are invariant across members of different subgroups (e.g., Carlton & Harris, 1992; Chen & Henning, 1985; Elder, 1996; Lawrence & Curley, 1989; Maller, 2001; Pae, 2004; Ryan & Bachman, 1992; Scheuneman & Gerritz, 1990; Shmitt & Dorans, 1990; Thissen et al., 1988, 1993). Ethnicity (e.g., Schmitt & Dorans, 1990), academic backgrounds (e.g., Pae, 2004), linguistic backgrounds (e.g., Chen & Henning, 1985; Ryan & Bachman, 1992), and disability status (e.g., Maller, 1997) have all been used to group people (e.g. Maller, 2001).

DIF analysis has been proposed as a tool for explaining group strengths and shortcomings; Differences in group strengths and weaknesses, on the other hand, rarely arise solely at the item level (Li & Suen, 2012). Li and Suen also explained that even though the item is the DIF's unit of analysis, it is not very useful for studying these differences.

Differential Skill Functioning (DSF) analysis, a variation of traditional DIF methods, has been proposed as an alternative. DSF occurs when examinees from different groups have different probabilities of successful performance in a specific sub-skill underlying the measured construct, despite having the same overall ability (Li & Suen, 2012). According to Li and Suen, although the logic and approach of DSF studies are similar to those of DIF analyses, the aim, or the unit of analysis, is a skill expressed by a number of items rather than a single item. Li and Suen (2012) extended the DIF approach to individual skill performance as evaluated by the Preliminary SAT/National Merit Scholarship Qualifying Test (PSAT/NMSQT) in order to compare aggregate groupings to the complete population matched on overall scores (Li & Suen, 2012).

According to McNamara and Roever's (2006) analyses of test and item bias, DIF and DSF are part of the overall work on test fairness. They help test users conduct a test that is fair to all participants and is not biased towards a certain group. If a test is biased, the inferences drawn from it are not based on genuine differences, and any measurement, inferences, or decisions drawn from it are useless. The validity of inferences can be jeopardized by a variety of variables common to language performance evaluations, such as raters, task, rating scale, candidate characteristics, and interactions among these (see McNamara, 1996 for a model and discussion); As a result, researchers and test users should be aware of test bias analysis, how it might affect test results, and how to identify and eliminate it.

Although it has been suggested in literature that language test performance may be influenced by background knowledge (e.g., Clapham, 1996; Hale, 1988), most DIF and DSF studies focus on comparisons between gender (e.g., Carlton & Harris, 1992), ethnicity (e.g., Schmitt & Dorans, 1990), or different language groups (e.g., Ryan & Bachman, 1992). According to Pae (2004), studies that looked at the impact of prior knowledge on test

performance are extremely rare. In this regard, DIF and DSF investigations for test-takers from different academic disciplines (e.g., Humanities vs. Sciences) would fill a gap in the DIF and DSF literature by revealing whether test items and skills function differently for examinees from different academic backgrounds but are matched to the primary construct that the test is intended to measure.

English Proficiency Test (EPT) is a test that is used to examine and assess the competency level of applicants who are interested in taking the Ph.D. exams. Given the importance of validity, test administrators of universities must guarantee that the test is not biased against a specific group. Since few studies have focused on academic background which creates DSF and DIF on different tests (Pae, 2004), test administrators may wonder whether students' academic backgrounds affect their performance or not. Thus, the purpose of the present study is to identify items and skills that exhibit DIF and DSF on EPT for examinees with different academic backgrounds (Humanities versus Sciences) using the Item Response Theory (IRT) and Mantel-Haenszel (MH) (Mantel & Haenszel, 1959) methodologies so that its administrators will be able to improve the validity of it for its applicants. This study examines differential functioning at both the item and skill levels, and if any disparities in item and skill performance are discovered between two groups, it attempts to see whether they can be improved.

2. Research Questions

According to the purpose of this study, the researchers try to answer the following questions:

1. Do items of English Proficiency Test (EPT) show Differential Skill Functioning (DSF) in favor of a particular academic discipline group using Mantel-Haenszel (MH) and Item Response Theory (IRT) procedures?
2. Do items of English Proficiency Test (EPT) show Differential Item Functioning (DIF) in favor of a particular academic discipline group using Mantel-Haenszel (MH) and Item Response Theory (IRT) procedures?
3. Do Item Response Theory (IRT) and Mantel-Haenszel (MH) detect similar Differential Item Functioning (DIF) and Differential Skill Functioning (DSF)?
4. What does the content analysis of DIF items reveal in terms of linguistic differences?

3. Method

3.1. Participants and Setting

The participants of the present study are 642 post-graduate students with an age range of 24 to 38. They were selected by convenience sampling from universities in Tehran. They were going to pursue their PhD studies in a variety of academic disciplines including Humanities, Sciences, and Engineering. Passing EPT is a prerequisite for them to be allowed to sit for the Ph.D. exam. Based on the academic background, the sample is divided into a reference group of 468 participants with Sciences and Engineering background and a focal group of 174 participants with Humanities background. The Sciences and Engineering group comprises students of chemistry, physics, mathematics, biology, mechanical engineering, electrical engineering, and civil engineering; and the Humanities group consists of students of social sciences, law, political sciences, management, Persian literature, and foreign languages. There were both male and female test takers in the sample. Because some studies identified 250 test takers per group as the minimum for adequate power in the MH procedure, a sample size of 642 test takers was chosen (Güler & Penfield, 2009; Mazor et al., 1992; Paek & Wilson, 2011; Rogers & Swaminathan, 1993). When the magnitude of the DIF is large, the power found in simulation studies with smaller sample sizes (50 or 100 test takers per group) is only moderately acceptable (Fidalgo et al., 2004; Fidalgo et al., 2007).

3.2. Materials

3.2.1. EPT

English Proficiency Test (EPT) is a high-stakes test of English that is being developed and administered by English department of the university. It is a prerequisite for master's degree holders aiming at participating in Ph.D. exams. It includes three sections with a time limit of 75 minutes. The first section including structure and written expression questions consists of 30 items, the second part includes vocabulary questions with 30 items, and the last section consists of 60 items including reading comprehension questions.

3.3. Procedures

3.3.1. Data Collection Procedures

The purpose of this study is to examine academic discipline differences in EPT so that its administrator will be able to improve the validity of it for its applicants. This study is part of a large-scale study. As a result, a total of 642 test takers were used for the pursuit of the research questions and data was collected by administering EPT in October, 2016. According to Penfield (2001), comparisons are frequently done between a single reference group, which is the favored group of examinees, and the focal group, which is the disadvantaged group of examinees, while undertaking a Differential Item Functioning (DIF) analysis (Abbott, 2016). As a result, the study's researchers divided the participants into two groups based on their academic discipline: a reference group of 468 students with a Sciences and Engineering background and a focal group of 174 students with a Humanities background, in order to see if the academic discipline of students has an impact on their performance. In DIF and DSF investigations, it is advised that more than one method of DIF or DSF analysis can be used to obtain more reliable results (Camilli, 2006; Camilli & Shepard, 1994; Pae, 2012; Uiterwijk & Vallen, 2005). The current study uses two methodologies to support this claim: (a) the Mantel-Haenszel (MH) model, and (b) the Item Response Theory (IRT) model. This increased the results' dependability and allowed for a comparison of the degree of concordance between the two methodologies' outcomes. The data was analyzed using DIFAS and Winsteps applications in addition to IRT and MH methods.

3.3.2. Mantel-Haenszel Procedures

DIF detection can be done in a variety of ways. The accuracy of them is determined by the disparities in the groups' ability levels. The initial step in detecting DIF is to match examinees with similar skill levels (Allalouf & Abramzon, 2008). Matching is difficult when there are significant differences in ability. Large ability gaps can result in decreased agreement across various DIF detection systems when it comes to DIF items, for example (as was found by Hambleton & Rogers, 1989).

The Mantel-Haenszel (MH) DIF detection method is a widely utilized method (see Holland & Thayer, 1988). Because it has been widely used in a number of studies (e.g., Azocar et al., 2001; Roznowski & Reith, 1999; Scheuneman & Gerritz, 1990; Schmitt et al., 1993), we utilized the Mantel Haenszel chi-square (MH- χ^2) approach to detect DIF (Mantel

& Haenszel, 1959). The MH technique was chosen for this investigation because of its simplicity, consistency across populations, and ease of interpretation.

When using the MH procedure to detect DIF, examinees are first grouped according to an estimate of ability, usually the total test score, and then a two-by-two contingency table is created for each level of ability, crossing group membership (reference and focal) and item performance (correct and incorrect) (Penfield, 2001).

Finally, in terms of academic discipline directedness, EPT was evaluated for the presence of DIF. Mantel Chi-square statistic (Mantel, 1963; Zwick et al., 1993; Zwick et al., 1997) and Liu-Agresti Cumulative Common Log-Odds Ratio (Liu & Agresti, 1996; Penfield & Algina, 2003) were obtained and analyzed for presence and direction of DIF in favor of reference or focal group using Differential Item Functioning Analysis System (DIFAS) program version 5.0.

3.3.3. Item Response Theory

An alternative approach to the elimination of items could be to investigate if the items that exhibit DIF genuinely assess the same construct across conditions even if they do so using a different method (Makransky & Glas, 2013). Such disparities can be modeled using group-specific item characteristics in Item Response Theory (IRT). This method is only valid if it can be demonstrated that the responses to the items in both groups are related to the same latent variable. To put it another way, the construct being measured must be the same in both groups. This may be demonstrated by examining if the same IRT model applies to the complete set of response data. This technique is justified by the fact that objects can have slightly different true parameters in different contexts. When statistical evidence supports the premise that the items assess the same construct across conditions, these differences can be modeled (Makransky & Glas, 2013). The study used BILOG-item MG's response theory (IRT)-based DIF detection approach, which compares latent trait item difficulty metrics across groups (Geranpayeh & Kunnan, 2007).

The present study was also carried out in the framework of IRT using Winsteps program. The basic premise of IRT is that each test item is defined by one or more

parameters, whereas each test taker is defined by a single ability parameter. The chance that a given test-taker will properly answer a given question is a function of both the item's and the test taker's factors. The response to one item is independent of the responses to other items if those requirements are met (Makransky & Glas, 2013).

4. Results

4.1. Results of Question One

The data was analyzed using DIFAS program to see whether DIF and DSF are present in items of English Proficiency Test across different academic disciplines. With this aim, the participants of Sciences and Engineering disciplines were taken as reference group (468 members) and the participants of Humanities discipline were taken as focal group (174 members). According to Penfield (2003), critical values of statistics of MH CHI and BD columns are 3.84 for a Type I error rate of 0.05 and 6.63 for a Type I error rate of 0.01 and a value greater than 2.0 or less than -2.0 in LOR Z column may be considered as an evidence of the presence of DIF. MH LOR column also has been explored to identify the direction of possible DIF in scale items. Positive values indicate DIF in favor of the reference group (Sciences and Engineering disciplines in this study), and negative values indicate DIF in favor of the focal groups (Humanities discipline). CDR flags any item for which the Mantel-Haenszel chi-square or the Breslow-Day chi-square statistic is significant at a Type I error rate of 0.025, and items that demonstrate DIF are divided into three groups at the Educational Testing Services (ETS) in the United States: Items with a negligible DIF are in Category A, items with an intermediate DIF are in Category B, and items with a large DIF are in Category C (Dodeen & Johanson, 2003). Tables 1, 2, and 3 display flagged items detected by Mantel-Haenszel theory using DIFAS program.

4.1.1. Grammar Skill Using DIFAS Program

As shown in Tables 1 and 2, items of grammar skill are flagged as DIF. Item 23 (MH CHI= 5.95, MH LOR= 0.49, LOR Z= 2.51 and BD= 0.00) has been categorized as having moderate (B) level of DIF and item 26 (MH CHI= 0.09, MH LOR= 0.07, LOR Z= 0.40 and BD= 5.12) has been categorized as having small (A) level of DIF. The value of MH CHI for item 23 is

more than 3.84 and the value of BD for item 26 is more than 3.84 and both of them show DIF toward reference group. In order to prove the existence of DIF items in the MH CHI column, LOR Z column was also studied and it demonstrates item 23 with DIF, because the value of it is more than +2.

Table 1

DIF Statistics: Dichotomous Items of Grammar Skill

NAME	MH CHI	MH LOR	LOR SE	LOR Z	BD	CDR	ETS
Q23	5.9533	0.4905	0.1952	2.5128	0.001	Flag	B
Q26	0.0969	0.0751	0.1865	0.4027	5.126	Flag	A

4.1.2. Vocabulary Skill Using DIFAS Program

In Tables 2 and 3, items of vocabulary skill are presented as showing DIF.

Table 2

DIF Statistics: Dichotomous Items of Vocabulary Skill

NAME	MH CHI	MH LOR	LOR SE	LOR Z	BD	CDR	ETS
Q38	0.3394	-0.1597	0.2337	-0.6834	9.957	Flag	A
Q41	6.4622	-0.5441	0.2082	-2.6134	0.983	Flag	B
Q49	0.9555	0.1962	0.1855	1.0577	6.136	Flag	A

Item 38 has been categorized as having small (A) level of DIF (MH CHI= 0.33, MH LOR= -0.15, LOR Z= -0.68 and BD= 9.95), item 41 (MH CHI= 6.46, MH LOR= -0.54, LOR Z= -2.61 and BD= 0.98) has been categorized as having moderate (B) level of DIF and Item 49 (MH CHI= 0.95, MH LOR= 0.19, LOR Z= 1.05 and BD= 6.13) has been categorized as having small (A) level of DIF. The value of BD for items 38 and 49 is more than 3.84 and the value of MH CHI for item 41 is more than 3.84. MH LOR column was studied to identify the direction of possible DIF in scale items. Items 38 and 41 show DIF toward focal group, because they have negative values and item 49 shows DIF toward reference group because it has positive value. In order to prove the existence of DIF items in the MH CHI column, LOR

Z column was also studied and it shows that item 41 shows DIF, because its value is less than -2.

4.1.3. Reading Comprehension Skill Using DIFAS Program

Out of 60 items of reading, 5 items are flagged as DIF. Item 64 (MH CHI= 8.20, MH LOR= -0.55, LOR Z= -2.90 and BD= 4.64) has been categorized as having moderate (B) level of DIF, item 68 (MH CHI= 1.54, MH LOR= 0.25, LOR Z= 1.32 and BD= 6.47) has been categorized as having small (A) level of DIF, item 72 (MH CHI= 0.24, MH LOR= 0.13, LOR Z= 0.60 and BD= 7.73) has been categorized as having small (A) level of DIF, item 86 (MH CHI= 0.55, MH LOR= -0.21, LOR Z= -0.87 and BD= 5.84) has been categorized as having small (A) level of DIF and item 90 (MH CHI= 11.51, MH LOR= -0.65, LOR Z= -3.50 and BD= 0) has been categorized as having moderate (B) level of DIF. As shown in Table 3, the value of MH CHI for items 64 and 90 is more than 3.84 and the value of BD for items 68, 72, and 86 is more than 3.84. In MH LOR column, items 64, 86, and 90 have negative values and show DIF toward focal group but items 68 and 72 have positive values and show DIF toward reference group. In order to prove the existence of DIF items in the MH CHI column, LOR Z column was also studied and it shows item 41 with DIF, because the value of it is less than -2. LOR Z column shows 2 items with DIF, both of them were less than -2.

Table 3

DIF Statistics: Dichotomous Items of Reading Skill

NAME	MH CHI	MH LOR	LOR SE	LOR Z	BD	CDR	ETS
Q64	8.2091	-0.5598	0.1927	-2.905	4.646	Flag	B
Q68	1.5466	0.2574	0.1946	1.3227	6.471	Flag	A
Q72	0.2468	0.1394	0.2308	0.604	7.732	Flag	A
Q86	0.5549	-0.2146	0.2465	-0.8706	5.841	Flag	A
Q90	11.5106	-0.6513	0.1859	-3.5035	0	Flag	B

Looking back at Tables 1, 2, and 3 and considering that 2 items of grammar skill (items 23 and 26), 3 items of vocabulary skill (items 38, 41, and 49) and 5 items of reading

skill (items 64, 68, 72, 86, and 90) have been flagged as DIF, it is noticed that although the number of DIF items in reading section is more than other skills, this skill does not show DSF, because out of 40 items belonging to reading skill, just 5 items have been flagged as DIF and three of these items indicate that focal group (Humanities academic discipline) is advantaged and two of them indicate that reference group (Sciences and Engineering academic disciplines) is favored. Therefore, it is concluded that items of EPT do not show DSF in favor of a particular academic discipline group using Mantel-Haenszel (MH) procedure.

In order to assess the presence of DIF and DSF by implementing IRT method, Winsteps was used. Two DIF detection methods are used in Winsteps (a Rasch analysis software program): (1) the Welch t-test and (2) the Mantel-Haenszel method (Linacre, 2010). The outcomes of these two methods are used to classify items using the ETS DIF Categories. The interest group is the focal group, whereas the reference group serves as a comparison group (Camilli & Shepard, 1994; Dorans & Holland, 1993). For this study, the participants of Sciences and Engineering disciplines were taken as the reference group (468 members) and the participants of Humanities discipline were taken as the focal group (174 members). Generally, the reference groups are expected to have the perceived advantage while the focal groups have the disadvantage.

Person Class (PC) is the indicator of the group which the participants belong to. PC1 belongs to Sciences and Engineering students (reference group) and PC2 belongs to Humanities students (focal group). DIF measure is the difficulty of item in person class. The hypothesis in the Welch probability column is that each item has the same difficulty for both the reference and focal groups, and it depicts the likelihood of observing the amount of DIF contrast by chance. As a result, values less than 0.05 indicate DIF. Welch t is a Welch's t-statistic that calculates the DIF significance (Runnels, 2013). Both the amount of the difference in logit units between the groups and the statistical significance of the difference should be considered when performing DIF Analysis in a Rasch context (Linacre, 2010). The DIF Contrast measure, according to Runnels (2013), depicts the difference between the item measures, or the difference in item difficulty between the two groups (a difference of at least

0.5 logits is required for the DIF to be noticeable). The DIF Contrast statistic represents the amount and direction of the difference in difficulty estimations between the reference and focal groups. Negative values imply that the item is more difficult for the focal group, whereas positive values suggest that the item is more difficult for the reference group (Runnels, 2013).

The ETS DIF Categories have been in operation for over 25 years and are assigned based on the direction, size, and relevance of the DIF statistics (Linacre, 2010). To “avoid recognizing items that demonstrate practically insignificant but statistically significant DIF,” the three categories were devised (Clauser & Mazor, 1998). All items are sorted into the DIF categories as described in Table 4 (Zwick, 2012).

Table 4

ETS Differential Item Functioning Categories from WINSTEPS

ETS DIF Category	Requirements
Class A (Negligible DIF)	$ \text{DIF Contrast} < 0.43$
Class B (Slight to Moderate DIF)	Doesn't meet the requirements for A or C
Class C (Moderate to Large DIF)	$ \text{DIF Contrast} > 0.64$

Tables 5, 6, and 7 demonstrate the results of IRT analysis for different academic discipline groups and their performance on different skills of EPT. Note that, due to space limitation, only DIF-flagged items are presented.

4.1.4. Grammar Skill Using Winsteps Program

In Table 5, 3 items are presented as showing DIF detected by IRT method. They are items 8, 20, and 23. Item 8 (DIF contrast= -0.38, Welch t= -2.04 and Welch prob. =0.04) and item 23 (DIF contrast= -0.47, Welch t= -2.51 and Welch prob. =0.01) show DIF in favor of Sciences and Engineering groups and item 20 (DIF contrast= 0.44, Welch t= 2.28 and Welch prob. = 0.02) shows DIF in favor of Humanities group. The value of item 8 in DIF contrast column is less than 0.43; therefore, it has been categorized as having small (A) level of DIF. In items 20

and 23, the value of DIF contrast is more than 0.43 and less than 0.64, therefore they show moderate (B) DIF.

4.1.5. Vocabulary Skill Using Winsteps Program

As Table 6 demonstrates, 3 items of vocabulary skill have been flagged as DIF. They are item 39 (DIF contrast= 0.36, Welch t= 1.91 and Welch prob. = 0.05), item 41 (DIF contrast= 0.39, Welch t= 1.93 and Welch prob. = 0.05), and item 58 (DIF contrast= 0.61, Welch t= 2.44 and Welch prob. =0.01).

All DIF items have positive values in DIF contrast column and show DIF toward focal group (Humanities students). Items 39 and 41 have a value less than 0.43 in DIF contrast column and they have been categorized as having a small (A) level of DIF but the value of DIF contrast for item 58 is more than 0.43 and less than 0.64, so it has been categorized as having a moderate (B) level of DIF.

Table 5

DIF Flagged Items of Grammar Skill Detected by IRT Model

Name	PC1 DIF Measure	PC2 DIF Measure	DIF Contrast	Welch t	Welch prob.
Q8	-1.41	-1.04	-0.38	-2.04	0.0421
Q20	0.46	0.02	0.44	2.28	0.0233
Q23	-0.35	0.12	-0.47	-2.51	0.0126

Table 6

DIF Flagged Items of Vocabulary Skill Detected by IRT Model

Name	PC1 DIF Measure	PC2 DIF Measure	DIF Contrast	Welch t	Welch prob.
Q39	0.23	-0.13	0.36	1.91	0.0565
Q41	0.68	0.29	0.39	1.93	0.0547
Q58	1.54	0.93	0.61	2.44	0.0152

4.1.6. Reading Comprehension Skill Using Winsteps Program

As shown in Table 7, two items of reading skill display DIF. Item 64 (DIF contrast= 0.48, Welch t= 2.55 and Welch prob. = 0.01) and item 90 (DIF contrast= 0.58, Welch t= 3.21 and Welch prob. = 0.00) have positive values in DIF contrast column and show DIF toward the Humanities group. Both of them show greater values than 0.43 and fewer values than 0.64 in DIF contrast column. Therefore, they have been categorized as having a moderate (B) level of DIF.

Referring back to Tables 5, 6, and 7 and considering that three items of grammar skill (items 8, 20, and 23), three items of vocabulary skill (items 39, 41, and 58) and two items of reading skill (items 64 and 90) have been flagged as DIF, it is noticed that the number of DIF detected items in grammar and vocabulary sections is more than reading section; but because out of 30 items, only 3 items showed DIF in these sections, it is concluded that items of EPT do not show DSF in favor of a particular academic discipline group using IRT procedure.

Table 7

DIF Flagged Items of Reading Skill Detected by IRT Model

Name	PC1 DIF Measure	PC2 DIF Measure	DIF Contrast	Welch t	Welch prob.
Q64	-0.96	-1.44	0.48	2.55	0.0112
Q90	-0.15	-0.73	0.58	3.21	0.0015

4.2. Results of Question Two

To investigate DIF in this study across academic disciplines, the data was analyzed by DIFAS program. With this aim, the participants of Sciences and Engineering disciplines were taken as the reference group and the participants of Humanities discipline were taken as the focal group. As shown in Table 8, of all 100 items, ten items show DIF.

DIF detected items include item 23 (MH CHI= 5.95, BD= 0.00), item 26 (MH CHI= 0.09, BD= 5.12), item 38 (MH CHI= 0.33, BD= 9.95), item 41 (MH CHI= 6.46, BD= 0.98),

item 49 (MH CHI= 0.95, BD= 6.13), item 64 (MH CHI= 8.20, BD= 4.64), item 68 (MH CHI= 1.54, BD= 6.47), item 72 (MH CHI= 0.24, BD= 7.73), item 86 (MH CHI= 0.55, BD= 5.84), and item 90 (MH CHI= 11.51, BD= 0). The value of MH CHI for items 23, 41, 64, and 90 is more than 3.84 and the value of BD for items 26, 38, 49, 64, 68, and 86 is more than 3.84.

These items are from different skills. Items 38, 41, 64, 86, and 90 in MH LOR column have negative values, so they demonstrate DIF in favor of the focal group and items 23, 26, 49, 68, and 72 have positive values and manifest DIF in favor of the reference group. Items 23, 41, 64, and 90 have been categorized as having a moderate (B) level of DIF and items 26, 38, 49, 68, 72 and 86 have been categorized as having a small (A) level of DIF. In order to prove the existence of DIF items in the MH CHI column, LOR Z column is also studied and it shows 4 items with DIF, one of them (item 23) is more than +2 and three of them (items 41, 64, and 90) are less than -2. As shown in Table 8, five positive and five negative values in MH LOR column indicated a good balance between the reference and focal groups directed items.

Table 8

DIF Detected Items of EPT Using Mantel-Haenszel Procedure

NAME	MH CHI	MH LOR	LOR SE	LOR Z	BD	CDR	ETS
Q23	5.9533	0.4905	0.1952	2.5128	0.001	Flag	B
Q26	0.0969	0.0751	0.1865	0.4027	5.126	Flag	A
Q38	0.3394	-0.1597	0.2337	-0.6834	9.957	Flag	A
Q41	6.4622	-0.5441	0.2082	-2.6134	0.983	Flag	B
Q49	0.9555	0.1962	0.1855	1.0577	6.136	Flag	A
Q64	8.2091	-0.5598	0.1927	-2.905	4.646	Flag	B
Q68	1.5466	0.2574	0.1946	1.3227	6.471	Flag	A
Q72	0.2468	0.1394	0.2308	0.604	7.732	Flag	A
Q86	0.5549	-0.2146	0.2465	-0.8706	5.841	Flag	A
Q90	11.5106	-0.6513	0.1859	-3.5035	0	Flag	B

Table 9 demonstrates the results of IRT analysis for different academic discipline groups and their performance on different items of EPT by Winsteps program. For simplicity, only items displaying DIF are presented.

As shown in Table 9, of all 100 items, 8 items show DIF. DIF detected items include items 8, 20, 23, 39, 41, 58, 64, and 90. These items are from different sections of EPT. Of all these items, items 8 and 23 have negative values in DIF contrast column and they show DIF in favor of Sciences and Engineering groups. Items 20, 39, 41, 58, 64, and 90 have positive value, so they manifest DIF in favor of Humanities group. Among these items, the value of items 8, 39, and 41 in DIF contrast column is less than 0.43; therefore, they show small (A) DIF. The value of DIF contrast for other items is more than 0.43 and less than 0.64; so according to Table 4.4, they show moderate (B) DIF.

Table 9

DIF Detected Items of EPT Using IRT Procedure

Name	PC1 DIF Measure	PC2 DIF Measure	DIF Contrast	Welch t	Welch prob.
Q8	-1.41	-1.04	-0.38	-2.04	0.0421
Q20	0.46	0.02	0.44	2.28	0.0233
Q23	-0.35	0.12	-0.47	-2.51	0.0126
Q39	0.23	-0.13	0.36	1.91	0.0565
Q41	0.68	0.29	0.39	1.93	0.0547
Q58	1.54	0.93	0.61	2.44	0.0152
Q64	-0.96	-1.44	0.48	2.55	0.0112
Q90	-0.15	-0.73	0.58	3.21	0.0015

4.3. Results of Question Three

Four items flagged as DIF by IRT, have been flagged by Mantel-Haenszel method using DIFAS program. They are items 23, 41, 64, and 90. Items 64 and 90 show moderate (B) DIF in favor of Humanities group, item 23 shows moderate (B) DIF and it advantaged the

reference group (Sciences and Engineering) by MH and IRT analyses using DIFAS and Winsteps, respectively; and finally, item 41 has been categorized as having small (A) DIF in favor of Humanities group using IRT procedure and having moderate (B) DIF in favor of Humanities group using MH procedure. Items 26, 38, 49, 68, 72, and 86 are DIF items detected only by MH method using DIFAS program. Among these items, items 49, 68, and 72 show small (A) DIF in favor of Sciences and Engineering groups; items 38 and 86, show small DIF in favor of Humanities group and item 26 has been categorized as having moderate (B) DIF in favor of Sciences and Engineering groups. Items 8, 20, 39, and 58 are items flagged as DIF by IRT procedure using Winsteps program.

Table 10

DIF Items Detected by Mantel-Haenszel and IRT

Item	DIF Size by MH/advantaged group		DIF Size by IRT/advantaged group	
Q8	-	-	A	Sciences & engineering
Q20	-	-	B	Humanities
Q23	B	Sciences & engineering	B	Sciences & engineering
Q26	B	Sciences & engineering	-	-
Q38	A	Humanities	-	-
Q39	-	-	A	Humanities
Q41	B	Humanities	A	Humanities
Q49	A	Sciences & engineering	-	-
Q58	-	-	B	Humanities
Q64	B	Humanities	B	Humanities
Q68	A	Sciences & engineering	-	-
Q72	A	Sciences & engineering	-	-
Q86	A	Humanities	-	-
Q90	B	Humanities	B	Humanities

Item 8 shows small (A) DIF in favor of Sciences and Engineering groups, items 20 and 58 have been categorized as having B DIF in favor of Humanities group and item 39 shows A DIF and advantaged the Humanities group. Table 10 illustrates the DIF size of items

and the group has been advantaged by each item; it also shows whether flagged items as DIF have been detected by both methods or not. In order to see whether there is any correlation between these two methods, SPSS is used and as shown in Table 10, $p < 0.05$ and it is proved that there is a positive correlation between MH and IRT methods.

Table 11

Correlation between IRT and MH Methods

		IRT	MH
IRT	Pearson Correlation	1	-.548*
	Sig. (2-tailed)		.043
	N	14	14
MH	Pearson Correlation	-.548*	1
	Sig. (2-tailed)	.043	
	N	14	14

**correlation is significant at the 0.05 level (2-tailed)*

4.4. Results of Question Four

In order to investigate whether DIF items are biased or not, the researchers used content analysis. Item 8 belongs to part one of grammar section. It is an incomplete sentence and students have to choose the one word that best completes the sentence.

Item 8. *Usually strokes happena combination of factors, such as certain medical conditions, inherited characteristics or unhealthy sounds.*

- A. *for*
- B. *of*
- C. *because of*
- D. *because*

The value of Welch probability for this item is 0.04 using IRT method. It is less than 0.05 and flagged as DIF. It is considered as having a small level of DIF because the value of DIF Contrast is -0.38 and according to Table 4, it is less than 0.43. Its negative value also indicates that the item favors Sciences and Engineering groups. Content analysis showed that this item has been related to Sciences groups and because of this, the Humanities group found it harder to choose the best word. Therefore, it can be concluded that DIF item constitutes bias and items function differently for two groups of test takers because they are biased.

Items 20, 23, and 26 belong to the grammar section related to written expressions. The participants have to identify the one underlined word phrase that must be changed in order for the sentence to be correct.

Item 20. The information on the Iranian panther is too limited since this animal is regarded
A B C D
as very shy in the wildlife.

This item detected as DIF, only by IRT method. The value of DIF contrast for this item is 0.44 and it is more than 0.43 and less than 0.64, so it demonstrates moderate DIF and positive value in DIF contrast column shows that this item is in favor of Humanities group, although the content of this item is more related to Biology courses. So, this DIF item does not demonstrate bias and it functions differently for two groups of test takers simply because they differ in their ability.

Item 23. Stars in our universe vary in temperature, color, bright, size, and mass.
A B C D

This item demonstrates DIF using MH and IRT methods by DIFAS and Winsteps program; but using DIFAS program, the value of MH CHI is 5.95 and it is more than 3.84, and ETS, flagged it as moderate DIF. The positive value in MH LOR column shows that it is in favor of Sciences and Engineering groups. Using IRT, it shows moderate DIF in favor of Sciences and Engineering groups (Welch prob. = 0.01 and DIF contrast= -0.47). Content analysis showed that this item is more related to Physics and Geology courses belonging to Sciences academic discipline and because it detected as DIF using both methods, it can be concluded that item 23 displays bias.

Item26. Dependence on the text messaging has spread rapid since mobile phones invaded
A B C D.
the market at the end of the last century.

The value of BD for this item is estimated as 5.12 by MH method using DIFAS software and it displays moderate amount of DIF. The positive value for MH LOR column indicates that this item has DIF in favor of Sciences and Engineering groups. Content analysis showed that this item is free from bias because it depends on no academic discipline and it functions differently, simply because two groups of test takers differ in their ability.

Items 38, 39 and 41 are related to part one of vocabulary skills. In these items, some words have been underlined. Four choices, marked (A), (B), (C), and (D), have been suggested as their synonyms and participants have to choose one of these choices that is the most appropriate synonym for the underlined word.

Item 38. *Presumably desolate desert and tundra areas actually harbor many forms of life.*

A. favor

B. halt

C. generate

D. shelter

This item flagged as DIF only by MH estimated by DIFAS program. The value of BD for this item is 9.957 and it is more than 3.84. According to ETS, it has been categorized as small DIF. The value of MH LOR is -0.15 and it shows that this item is in favor of Humanities group. It has been shown in content analysis that this item is more related to Geography courses belonging to Humanities academic discipline, so it constitutes bias.

Item 39. *Due to large deposits of debris left by glaciers, the Mid western plain of United States is an extremely fertile area for farming.*

A. level

B. rich

C. spacious

D. arid

The value of Welch prob. for this item is 0.05 using IRT method, so it is flagged as a DIF item. According to Table 4, the value of 0.36 in DIF Contrast column indicates that this item has been categorized as having small DIF, because it is less than 0.43. The positive value of DIF Contrast for this item shows that it is in favor of Humanities group. It means that this item is hard for Sciences and Engineering groups. Content analysis showed that the content of this item is related to Geography and Humanities academic discipline, therefore, it can be concluded that item 39 constitutes bias and because of this, it was easier for the Humanities group.

Item 41. *Poorly constructed dwellings cannot withstand severe storms.*

A. Houses

B. Piers

C. Boats

D. roads

This item demonstrates DIF using MH and IRT method by DIFAS and Winsteps program, but the value of MH CHI is 6.46 using DIFAS program and it is more than 3.84, and ETS, flagged it as moderate DIF. The negative value in MH LOR column shows that it is in favor of Humanities group. Using IRT, it displays small DIF in favor of Humanities group (Welch prob. = 0.05 and DIF contrast= 0.39); however, content analysis showed that it is related to Engineering academic discipline. Therefore, it can be concluded that this item does not display bias and it functions differently just because the two groups differ in their ability.

Items 49 and 58 belong to part two of the vocabulary section, one word is removed from questions and four choices have been suggested for it in these items. The participants have to choose the best option.

Item 49. *A law that _____ tobacco advertising in newspapers and magazines has just been made public.*

A. prohibits

B. recycles

C. surrenders

D. overcomes

This item was detected as DIF only by MH method using DIFAS program. The amount of BD for this item is 6.13 and it is more than 3.84. The ETS column shows that this item has been categorized as having a small level of DIF. The value of MH LOR for this item is 0.19 and it indicates that the Sciences and Engineering group is advantaged; although the content analysis indicated that this item is more related to Law courses and Humanities academic discipline, it can be said that this DIF detected item does not display bias.

Item 58. *We can _____ the power of the wind to generate electricity.*

A. harness

B. justify

C. engender

D. obey

The value of Welch prob. for this item is 0.01 using IRT method. The value of DIF contrast for this item is 0.61 and it displays a moderate amount of DIF. Its positive value in DIF contrast columns indicates that the item favors Humanities group. It means that Sciences and Engineering group find it harder to choose the best option; however, the content of this item is more related to Engineering academic discipline. So it can be concluded that this item does not show bias.

Items 64, 68, 72, 86, and 90 belong to the reading comprehension section. These items are related to different passages. In these items, the participants have to read passages and answer the following questions. Item 64 belongs to passage one.

Item 64. *The word "vogue" in line 14 is closest in meaning to.....*

A. tendency

B. prevalence

C. challenge

D. marvel

Using DIFAS and Winsteps program, this item demonstrates DIF for both methods. But the value of MH CHI is 6.46 and it is more than 3.84 using DIFAS program. It is flagged as moderate DIF. The negative value in MH LOR column shows that it is in favor of Humanities group. Using IRT procedure with Winsteps program, it shows moderate DIF in favor of Humanities group (Welch prob. = 0.01 and DIF contrast= 0.48). Content analysis showed that the content of the first passage is about Arabic language and it is related to Humanities academic discipline and because of this, item 64 is easier for Humanities students rather than Sciences and Engineering students. Therefore, it can be concluded that this item constitutes bias. Items 68 and 72 are related to passage two.

Item 68. *What is the main point of the first paragraph?*

- A. The waves created by ocean currents are very large.*
- B. Despite the strength of the wind, it only moves surface water.*
- C. Deep ocean water is seldom affected by forces that move water.*
- D. The tides are the most powerful force to affect the movement of ocean water.*

The value of BD for this item is 6.47 using MH method by DIFAS and it is more than 3.84. This item is flagged as small DIF. The positive value of MH LOR for this item indicates that it is toward Sciences and Engineering group. It means that this item is hard for the Humanities group. Content analysis indicated that the passage to which this item belongs is more related to Physics and Geology belonging to Sciences academic discipline, so this item constitutes bias.

Item 72. *The word “configuration” in line 19 is closest in meaning to.....*

- A. unit*
- B. center*
- C. surface*
- D. arrangement*

The value of BD for this item is 7.73 using MH method by DIFAS and it is more than 3.84. This item is flagged as small DIF. The positive value of MH LOR for this item indicates that Sciences and Engineering group is advantaged. Analyzing the content of this item showed that the text to which this item belongs is more related to Physics and Geology belonging to Sciences academic discipline, so this item constitutes bias. Item 86 belongs to passage four.

Item 86. *The phrase “occupy the spotlight” in line 23 is closest in meaning to*

- A. receive the most attention*
- B. go the furthest*
- C. conquer territory*
- D. lighten the load*

Using MH method done by DIFAS program, the value of BD for this item is estimated as 5.84 and because it is more than 3.84, this item is flagged as DIF. The value of MH LOR for this item is -0.21 and according to ETS column, it has been categorized as having a small level of DIF. Content analysis indicated that this item must be in favor of the Sciences group because it is more related to Sciences courses but its negative value in MH LOR showed DIF in favor of the Humanities group. Therefore, it can be concluded that this item does not constitute bias. Item 90 belongs to passage five.

Item 90. *What aspect of drama does the author discuss in the first paragraph?*

- A. The reason drama is often unpredictable*
- B. The seasons in which dramas were performed*
- C. The connection between myths and dramatic plots*
- D. The importance of costumes in early drama*

This item has been flagged as DIF using MH and IRT methods by DIFAS and Winsteps software. Using DIFAS program, the value of MH CHI is 11.51 and it is more than 3.84, and ETS, flagged it as moderate DIF. The negative value in MH LOR column shows that it is in favor of the Humanities group. Using IRT and MH procedures with Winsteps program, we can show moderate DIF which indicates that Humanities group is advantaged (Welch prob. = 0.00 and DIF contrast= 0.58). Content analysis indicated that the passage to which this item belongs is more related to Literature belonging to Humanities, so this item displays bias.

5. Discussion

In this study, we examined the EPT for DIF attributable to academic disciplines. A two-step approach was used: First, the test items were examined for DIF, and second, the items that were flagged were subjected to content analysis by the researchers. In following sections, the overall findings of DIF/ DSF analysis are discussed.

Out of 100 items, 10 items were flagged as DIF in the statistical analysis. Among these, 2 items belonged to the grammar section, 3 items belonged to the vocabulary section, and 5 items belonged to the reading sections. The number of DIF items in the reading part is

more than the two other parts, but because only 5 items out of 40 were flagged as DIF (12.5%) and out of these 5 items, three of them indicated that the Humanities group is advantaged and two of them indicated that the Sciences and Engineering group is advantaged, it is concluded that items of EPT do not show DSF in favor of a particular academic discipline by using MH procedure.

As shown in Tables 5, 6, and 7, three items of grammar skill (10%), three items of vocabulary skill (10%), and two items of reading skill (5%) have been flagged as DIF. Among these three DIF detected items belonging to grammar skill, 2 items advantaged Sciences and Engineering group and one item advantaged Humanities group. Hence, 10% of grammar items show DIF and these items were not in favor of a particular academic discipline and cancel out each other, it can be concluded that grammar skill does not show DSF. Although all DIF detected items of vocabulary skill advantaged Humanities group, 3 items out of 30 are too small to consider this skill as having DSF. Several researchers examined the potential effects of DIF and DSF cancellation at test level in this regard (See e.g., Drasgow, 1987; Maller, 2001; Nandakumar, 1993; Roznowski, 1987; Roznowski & Reith, 1999). Investigating research question one, it may suggest that there existed some cancellation effects at test level because the expected score differences for the Humanities and Sciences were rather small in comparison to the number of items identified as indicating DIF for or against each academic category (Drasgow, 1987).

Investigating research question two, ten out of 100 items of EPT (10 percent) exhibited DIF by academic discipline. Out of these, fifty percent showed DIF in favor of the Humanities group and fifty percent showed DIF in favor of the Sciences and Humanities groups. Five items exhibited moderate (B) DIF and the others showed small (A) DIF. Therefore, it is concluded that the items of EPT show DIF in favor of a particular academic discipline group using Mantel-Haenszel procedure. However, previous research suggests that approximately one-third to one-half of personality test items show potential DIF. Waller et al. (2000) discovered potential DIF by race (Black vs. Caucasian) on 38% of the MMPI questions, while Sheppard et al. (2006) discovered potential DIF by race/culture (Chinese vs.

American) on 58% of the items in the Revised NEO Personality Inventory's conscientiousness scale (NEO-PI-R).

IRT identified 8 out of 100 items with DIF. Among these, twenty five percent indicated DIF in favor of the Sciences and Engineering group and seventy five percent advantaged the Humanities group. Five of these DIF detected items (62.5%) exhibited moderate (B) DIF and three of them (37.5%) exhibited small (A) DIF. As a result, it can be concluded that the EPT which was the object of this investigation shows DIF toward a particular group included in this study using IRT procedure. Since no studies to date have identified DIF in the EPT, the findings of this study have limitations in comparison across studies and require further investigations. For example, reviewing these items by gender and academic discipline as well as querying the students qualitatively may assist in understanding the differences between the group performances; However, as the first DIF study on the EPT, it is important to describe for future explanation that there were fourteen DIF items in the EPT across academic disciplines in the context of universities in Tehran. Thus, the revision or elimination of DIF items should be done with caution because they can represent important content for the future test construction. Using SPSS, it was found that there was a positive correlation between MH and IRT procedures.

For test development, DIF analysis is a useful tool. When significant amounts of DIF are found and eliminated from tests, they are more likely to be valid and less biased. According to the results of the content analysis across the two subscales, items dealing with science-related issues were differentially easier for the Sciences, whereas items addressing human interactions were differentially easier for the Humanities. Items 8 and 39 showed small (A) DIF using IRT method. Item 8 advantaged Sciences and Engineering group and item 39 advantaged Humanities group. They are also detected as DIF items by content analysis. Items 38, 68, and 72 were detected as small (A) DIF using MH procedure. The first one was in favor of Humanities group and the last two items belonging to passage two of the reading section, indicated DIF in favor of the Sciences and Engineering group. The content of the text is about tides and waves in the sea and the ocean which is more related to Physics and Geology belonging to Sciences academic discipline, so this item shows bias. Items 23,

64, and 90 exhibited moderate (B) DIF using IRT and MH procedures. Item 23 advantaged Sciences and Engineering group and the last two items belonging to passage one and passage five of the reading section, advantaged Humanities group. The content of the passage one is about Arabic language and the content of passage five is about drama. Hence, both of them are related to Humanities courses, it is concluded that these items display bias and it is better to modify them. This pattern contrasts with findings from gender DIF research, which found that females performed better than a matched group of males on reading items pertaining to human connections, but males were more likely to outperform a comparable group of females on science-related issues (e.g., Carlton & Harris, 1992; Curley & Schmitt, 1993; Lawrence & Curley, 1989; Lawrence et al., 1988; Maller, 2001; O'Neill & McPeck, 1993; O'Neill et al., 1993; Scheuneman & Gerritz, 1990).

However, some items were not consistent with the DIF pattern, which may limit the generalization of the findings. Items 20 and 58 indicated moderate (B) DIF in favor of the Humanities group using IRT method, although content analysis showed that these items are more related to Sciences courses and it is better to say that they are in favor of the Sciences group. Items 49 and 86 detected as small (A) DIF using MH procedure. The result of DIF procedures is not consistent with the result of content analysis. The content of item 49 is about law belonging to the Humanities group, but using MH, this item advantaged the Sciences group. Item 86 belongs to passage four which is related to diseases and medicine, so it is better to say that this item is in favor of the Sciences group but using MH, it has been shown that this item advantaged the Humanities group. Item 41 was detected as moderate (B) DIF and small (A) DIF using MH and IRT, respectively. Using DIF procedures, this item advantaged the Humanities group, although content analysis showed that the text of this item is related to Engineering academic discipline. Content analysis of item 26 indicated that this item is not in favor of a particular academic discipline but it was detected as moderate (B) DIF in favor of Sciences and Engineering group by using MH procedure. As numerous studies have argued, DIF items that deviate from the conventional DIF pattern demonstrate that content analysis alone is not a reliable means of detecting DIF (e.g., Engelhard et al., 1990; Plake, 1980; Rengel, 1986; Sandoval & Miille, 1980). However, a fair evaluation is necessary to make judgmental analyses of items that stereotype and under-represent a

particular population of examinees, because those items may cause the subgroup to be less motivated to do well on the test (Clauser & Mazor, 1998; Tittle, 1982).

In a similar vein, Scheuneman and Gerritz (1990) found that item content alone is not a reliable predictor of DIF direction, implying the inclusion of other unknown elements that lead to DIF. A large DIF value indicates the presence of additional structures that may function differently across the reference and focal groups because DIF approaches are based on comparable examinees.

6. Conclusion

This study investigated academic discipline differences in the English Proficiency Test (EPT) among 642 post-graduate students who were going to pursue their Ph.D. studies in a variety of academic disciplines including Humanities, Sciences, and Engineering. The results revealed that there were fourteen DIF items (4 grammar items, 5 vocabulary items, and 5 reading comprehension items) in the EPT across academic disciplines with Humanities group reporting a higher grade than Sciences and Engineering groups using Mantel-Haenszel (MH) and Item Response Theory (IRT) procedures. Four items (items 23, 41, 64, and 90) have been flagged as DIF by MH and IRT procedures, 6 items detected as DIF using only MH procedure (items 26, 38, 49, 68, 72, and 86) and 4 items have been flagged as DIF items only by IRT method (items 8, 20, 39, and 58). Six items have been categorized as having moderate (B) DIF (items 20, 23, 26, 58, 64, and 90), seven items have been categorized as having small (A) DIF (items 8, 38, 39, 49, 68, 72, and 86) and one of them which is flagged as DIF by both procedures, showed small DIF by IRT and moderate DIF by MH method (item 41). Winsteps and DIFAS software were utilized in this study to detect DIF using IRT and MH procedures, respectively. This study showed that the items of EPT do not show DSF in favor of a particular academic discipline using both MH and IRT methods.

SPSS was used to determine Mantel-Haenszel and Item response theory procedures correlate with each other and the result showed that there is a positive correlation between IRT and MH.

To see whether DIF detected items are biased or not, the researchers did content analysis. Interpreting DIF items in educational assessment is not an easy task, and as our attempt to find an explanation in the content analysis illustrates, it is even more difficult in the context of second language testing. Nonetheless, because test content and the types of tasks posed by particular test items are invariably major determinants of test performance, we believe that a closer analysis of the relationships between item content and DIF is one way for conducting future research. The result of the content analysis showed that out of 14 items, 8 items (items 8, 23, 38, 39, 64, 68, 72, and 90) are biased and 3 of these items were flagged as DIF using MH and IRT methods. Content analysis showed that 6 items are not biased; the content of item 26 is not related to a particular academic discipline, so it is free from bias, but the content analysis of 5 other items showed a different advantaged group from the one determined by the statistical procedures. So it cannot be explained precisely whether these items constitute bias or not, but it can be concluded that they are not biased and they function differently because the two groups differ in their abilities.

To summarize, the goal of a DIF analysis is to guarantee that test equity or fairness is met. The statistical identification of items that show signs of DIF is a critical step in achieving this goal. Because tests are inherently multidimensional, and multidimensionality is the main cause of DIF, a better understanding of test dimensionality and the effects of these dimensions on DIF could lead to a more accurate test score interpretation, greater control over the influence of relevant auxiliary dimensions, and a reduction in the influence of unintended and irrelevant nuisance dimensions. Using the DIF detection approach, this study has made a first contribution to the issue of test fairness for all test takers, regardless of their academic subject. Although it needs to be replicated with different samples, preliminary content analysis of those items flagged for DIF reveals that DIF may be connected with content characteristics particular to group membership (e.g., science-related topics or human issues). It should be emphasized that the preliminary content analysis described in this study was based on the researchers' subjective opinion rather than that of a panel of content experts, and thus could be subject to misclassification. To make the measurement procedure as precise as feasible, this study identifies the need for regular evaluation and revision of tests. The researchers' main recommendation is for test developers and designers. Test

developers play critical roles in the assessment process. They should be aware of the negative effects of DIF and DSF on test validity and test results so as to design and construct tests which do not advantage any groups by considering as many factors as possible. The results of this study offer test developers and designers to develop tests free from bias and pay attention to choice of test topics so that as far as possible no academic discipline group is clearly disadvantaged. It is also recommended that the items displaying significant DIF be analyzed and revised.

Conflict of interest

The author(s) certify/certifies that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in the present research paper.

References

- Abbott, M. L. (2016). A confirmatory approach to differential item functioning on an ESL reading assessment. *Language Testing*, 24(1), 7-36. DOI: 10.1177/0265532207071510.
- Abedi, J., Bailey, A.L., Butler, F., Castellon-Wellington, M., Leon, S., & Mirocha, J. (2005). *The Validity of Administering Large-Scale Content Assessments to English Language Learners: An Investigation from Three Perspectives* (CSE Report 663). Retrieved from <http://cse.ucla.edu/products/reports/r663.pdf#page=87>.
- Allalouf, A., & Abramzon, A. (2008). Constructing better second language assessments based on differential item functioning analysis. *Language Assessment Quarterly*, 5(2), 120-141. DOI: <https://doi.org/10.1177/026553228500200204>.
- Azocar, F., Arean, P., Miranda, J., & Muñoz, R. F. (2001). Differential item functioning in a Spanish translation of the Beck Depression Inventory. *Journal of Clinical Psychology*, 57(3), 355-365. DOI: 10.1002/jclp.1017.
- Bachman, L. F. (2004). *Statistical analyses for language testing*: Cambridge, UK: Cambridge University Press.
- Camilli, G. (2006). Test fairness. *Educational Measurement*, 4, 221-256. DOI: <https://doi.org/10.1002/jclp.1017>.
- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Newbury Park, CA: Sage.
- Carlton, S.T., & Harris, A.M. (1992). Characteristics associated with Differential Item Functioning on the Scholastic Aptitude Test: gender and majority /minority group comparisons (ETS Research Report, 92-64). Princeton, NJ: ETS Research Report. DOI: [doi/pdf/10.1002/j.2333-8504.1992.tb01495.x](https://doi.org/10.1002/j.2333-8504.1992.tb01495.x)
- Chen, Z., & Henning, G. (1985). Linguistic and cultural bias in language proficiency tests. *Language Testing*, 2(2), 155-163. DOI: <https://doi.org/10.1177/026553228500200204>

- Clapham, C. (1996). *The development of IELTS: A study of the effect of background on reading comprehension*. Cambridge: Cambridge University Press.
- Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice*, 17(1), 31-44. DOI: <https://doi.org/10.1111/j.1745-3992.1998.tb00619.x>
- Curley, W. E., & Schmitt, A. P. (1993). Revising SAT-Verbal items to eliminate Differential Item Functioning. New York: College Entrance Examination Board.
- De Ayala, R. J., Kim, S., Stapleton, L. M., & Dayton, C. M. (1999, July). *A reconceptualization of differential item functioning*. Paper Presented at the Annual Meeting of American Educational Research Association (AERA), Montreal, Canada.
- Dodeen, H., & Johanson, G. A. (2003). An Analysis of Sex-related Differential Item Functioning in Attitude Assessment. *Assessment & Evaluation in Higher Education*, 28(2), 129-134. DOI: 10.1080/02602930301667.
- Dorans, N., & Holland, P. W. (1993). DIF detection and description: Mantel–Haenszel and standardization. In P. W. Holland, & H. Wainer (Eds), *Differential Item Functioning* (pp. 35–66). Hillsdale, NJ: Lawrence Erlbaum. DOI: 10.21031/epod.368081.
- Drasgow, F. (1987). Study of the measurement bias of two standardized psychological tests. *Journal of Applied Psychology*, 72(1), 19.
- Elder, C. (1996). The effect of language background on “foreign” language test performance: The case of Chinese, Italian, and Modern Greek. *Language Learning*, 46(2), 233-282.
- Engelhard Jr, G., Hansche, L., & Rutledge, K. E. (1990). Accuracy of bias review judges in identifying differential item functioning on teacher certification tests. *Applied Measurement in Education*, 3(4), 347-360.
- Geranpayeh, A., & Kunnan, A. J. (2007). Differential Item Functioning in Terms of Age in the Certificate in Advanced English Examination. *Language Assessment Quarterly*, 4(2), 190-222. DOI: 10.1080/15434300701375758.

- Güler, N., & Penfield, R. D. (2009). A comparison of the logistic regression and contingency table methods for simultaneous detection of uniform and nonuniform DIF. *Journal of Educational Measurement*, 46(3), 314-329.
- Hale, G. A. (1988). Student major field and text content: interactive effects on reading comprehension in the Test of English as a Foreign Language. *Language Testing*, 5(1), 49-61.
- Hambleton, R. K., & Rogers, H. J. (1989). Detecting potentially biased test items: Comparison of IRT area and Mantel-Haenszel methods. *Applied Measurement in Education*, 2(4), 313-334.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum.
- Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Klieme, E., & Baumert, J. (2001). Identifying national cultures of mathematics education: Analysis of cognitive demands and differential item functioning in TIMSS. *European Journal of Psychology of Education*, 16(3), 385-402. DOI:10.1007/BF03173189.
- Koo, J., Becker, B. J., & Kim, Y. S. (2013). Examining differential item functioning trends for English language learners in a reading test: A meta-analytical approach. *Language Testing*, 31(1), 89-109. DOI: 10.1177/0265532213496097.
- Lawrence, I.M., & Curley, W.E. (1989). Differential Item Functioning for males and females on SAT-Verbal Reading subscore items: follow-up study (ETS Research Report 89-22). Princeton, NJ: Educational Testing Service.
- Lawrence, I.M., Curley, W.E., & McHale, F.J. (1988). Differential item functioning for males and females on SAT verbal reading subscore items (Report No. 88-4). New York: College Entrance Examination Board.

- Li, H., & Suen, H. K. (2012). Detecting native language group differences at the subskills level of reading: A differential skill functioning approach. *Language Testing*, 30(2), 273-298.
- Linacre, J. (2010). *Winsteps Rasch measurement: Computer software manual (version 3.70.0.2)*. Retrieved from <http://www.winsteps.com>
- Liu, I. M., & Agresti, A. (1996). Mantel-Haenszel-type inference for cumulative odds ratios with a stratified ordinal response. *Biometrics*, 52, 1223-1234.
- Makransky, G., & Glas, C. A. W. (2013). Modeling differential item functioning with group-specific item parameters: A computerized adaptive testing application. *Measurement*, 46(9), 3228-3237. DOI: 10.1016/j.measurement.2013.06.020.
- Maller, S. J. (1997). Deafness and WISC-III item difficulty: Invariance and fit. *Journal of School Psychology*, 35(3), 299-314.
- Maller, S. J. (2001). Differential item functioning in the WISC-III: Item parameters for boys and girls in the national standardization sample. *Educational and Psychological Measurement*, 61(5), 793-817.
- Mantel, N. (1963). Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association*, 58(303), 690-700.
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22(4), 719-748.
- Mazor, K. M., Clauser, B. E., & Hambleton, R. K. (1992). The effect of sample size on the functioning of the Mantel-Haenszel statistic. *Educational and Psychological Measurement*, 52(2), 443-451.
- McNamara, T. F. (1996). *Measuring second language performance*. London: Addison Wesley Longman.

- McNamara, T. F., & Roever, C. (2006). *Language testing: The social dimension*. Madison: John Wiley & Sons.
- Nandakumar, R. (1993). Simultaneous DIF Amplification and Cancellation: Shealy-Stout's Test for DIF. *Journal of Educational Measurement*, 30(4), 293-311.
- O'Neill, K. A., & McPeck, W. M. (1993). Item and test characteristics that are associated with differential item functioning. In P. W. Holland, & H. Wainer (Eds.), *Differential item functioning* (pp. 255-276). New Jersey: Lawrence Erlbaum Associates, Inc.
- O'Neill, K.A., McPeck, W.M., & Wild, C.L. (1993). *Differential Item Functioning on the Graduate Management Admission Test* (ETS Research Report 93-35). Princeton, NJ: ETS.
- Pae, T. I. (2004). DIF for examinees with different academic backgrounds. *Language Testing*, 21(1), 53-73.
- Pae, T. I. (2012). Causes of gender DIF on an EFL language test: A multiple-data analysis over nine years. *Language Testing*, 29(4), 533-554.
- Paek, I., & Wilson, M. (2011). Formulating the rasch differential item functioning model under the marginal maximum likelihood estimation context and its comparison with Mantel-Haenszel procedure in short test and small sample conditions. *Educational and Psychological Measurement*, 71(6), 1023-1046.
- Park, G. P., & French, B. F. (2013). Gender differences in the Foreign Language Classroom Anxiety Scale. *System*, 41(2), 462-471. DOI: 10.1016/j.system.2013.04.001.
- Penfield, R. D. (2001). Assessing Differential Item Functioning Among Multiple Groups: A Comparison of Three Mantel-Haenszel Procedures. *Applied Measurement in Education*, 14(3), 235-259. DOI: 10.1207/s15324818ame1403_3.
- Penfield, R. D. (2003). Applying the Breslow-Day test of trend in odds ratio heterogeneity to the analysis of nonuniform DIF. *Alberta Journal of Educational Research*, 49(3), 231-243.

- Penfield, R. D., & Algina, J. (2003). Applying the Liu-Agresti Estimator of the Cumulative Common Odds Ratio to DIF Detection in Polytomous Items. *Journal of Educational Measurement, 40*(4), 353-370. DOI: https://doi.org/10.1207/S15324818AME1403_3.
- Plake, B. S. (1980). A comparison of a statistical and subjective procedure to ascertain item validity: One step in the test validation process. *Educational and Psychological Measurement, 40*(2), 397-404.
- Rengel, E. (1986). Agreement between statistical and judgmental item bias methods. Retrieved from ERIC database. (ED289 890)
- Rogers, H. J., & Swaminathan, H. (1993). A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. *Applied Psychological Measurement, 17*(2), 105-116.
- Roznowski, M. (1987). Use of tests manifesting sex differences as measures of intelligence: Implications for measurement bias. *Journal of Applied Psychology, 72*(3), 480-483.
- Roznowski, M., & Reith, J. (1999). Examining the measurement quality of tests containing differentially functioning items: Do biased items result in poor measurement? *Educational and Psychological Measurement, 59*(2), 248-269.
- Runnels, J. (2013). Measuring differential item and test functioning across academic disciplines. *Language Testing in Asia, 3*(1), 12-29.
- Ryan, K. E., & Bachman, L. F. (1992). Differential item functioning on two tests of EFL proficiency. *Language Testing, 9*(1), 12-29.
- Sandoval, J., & Miille, M. W. (1980). Accuracy of judgments of WISC-R item difficulty for minority groups. *Journal of Consulting and Clinical Psychology, 48*(2), 249-253.
- Saville, N. (2005). Setting and monitoring professional standards: A QMS approach. *Cambridge ESOL Research Notes, 22*, 2-5.

- Scheuneman, J. D., & Gerritz, K. (1990). Using differential item functioning procedures to explore sources of item difficulty and group performance characteristics. *Journal of Educational Measurement*, 27(2), 109-131.
- Schmitt, A. P., & Dorans, N. J. (1990). Differential Item Functioning for Minority Examinees on the SAT. *Journal of Educational Measurement*, 27(1), 67-81.
- Schmitt, N., Hattrup, K., & Landis, R. S. (1993). Item bias indices based on total test score and job performance estimates of ability. *Personnel Psychology*, 46(3), 593-611.
- Sheppard, R., Han, K., Colarelli, S. M., Dai, G., & King, D. W. (2006). Differential item functioning by sex and race in the Hogan Personality Inventory. *Assessment*, 13(4), 442-453.
- Shohamy, E. (2000). Fairness in language testing. In A. J. Kunnan (Ed.), *Fairness and validation in language assessment* (pp. 15–19). Cambridge: Cambridge University Press.
- Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 147–169). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 67–113). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Tittle, C.K. (1982). Use of judgmental methods in item bias studies. In R.A. Berk (Ed.), *Handbook of methods of detecting test bias* (pp. 31–30). Baltimore, MD: Johns Hopkins University Press.
- Uiterwijk, H., & Vallen, T. (2005). Linguistic sources of item bias for second generation immigrants in Dutch tests. *Language Testing*, 22(2), 211-234.

- Waller, N. G., Thompson, J. S., & Wenk, E. (2000). Using IRT to separate measurement bias from true group differences on homogeneous and heterogeneous scales: an illustration with the MMPI. *Psychological Methods*, 5(1), 125-146.
- Yu, L., Lei, P. W., & Suen, H. K. (2006, April). *Using a differential item functioning (DIF) procedure to detect differences in opportunity to learn (OTL)*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, Colifornia.
- Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF)*. Ottawa, Canada: National Defense Headquarters.
- Zwick, R. (2012). A review of ETS differential item functioning assessment procedures: Flagging rules, minimum sample size requirements, and criterion refinement (Research Report ETS RR-12-08). Princeton, NJ: Educational Testing Service.
- Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement*, 30, 233–251.
- Zwick, R., Thayer, D. T., & Mazzeo, J. (1997). Descriptive and inferential procedures for assessing differential item functioning in polytomous items. *Applied Measurement in Education*, 10(4), 321-344.

The Relationship between Iranian EFL Learners' Critical Thinking and their Inferential Ability in Reading Comprehension

Azade Ahmadi Hafshejani¹; Sara Falahi²

1. Master of ELT, English Language and Literature Department, Literature and Humanities faculty, University of Sistan and Balouchestan, Zahedan, Iran.

Azadeahmadi2012@gmail.com (corresponding author)

2. Master of ELT, English Language and Linguistics Department, Literature and Humanities Faculty, University of Razi, Kermanshah, Iran.

sara.falahi90@gmail.com

Abstract

The present study made an attempt to see if there was any statistically significant relationship between Iranian English as a Foreign Language (EFL) learners' critical thinking and their inferential ability in reading comprehension. Moreover, it attempted to see if there was any statistically significant difference between Iranian EFL learners' critical thinking and their inferential ability in reading comprehension. The participants were 127 English Literature, English Language Teaching and English Translation students at University of Isfahan, Islamic Azad University (Khorasgan Branch), Sheikh Bahaie University, Payame Noor University, and Islamic Azad University (Najafabad Branch). A quick placement test, a critical research questionnaire, and an inferential reading comprehension test were administered to the participants. The results revealed that there was a positive significant correlation between the critical thinking and inferential ability in reading comprehension. Also, the observed Z score showed that there was no statistically significant difference between the two correlation coefficients regarding gender. The findings of this study emphasize the importance of the relationship between critical thinking, and inferential ability in reading comprehension regarding gender and language proficiency in language learning and teaching.

Keywords: Critical thinking, Inferential ability, Reading comprehension, Gender, Language proficiency, Iranian EFL learners

Citation: Ahmadi Hafshejani, A., Falahi, S. (2024). *Applications of Language Studies*, 2(2), 59-86.

Received: 2024-02-28

Revised: 2024-07-31

Accepted: 2024-08-06

Article type: Research paper

P-ISSN: 2588-6975

E-ISSN: 2588-6983

Publisher: Hazrat-e Masoumeh University Press

DOI:10.22034/jals.2024.2023703.1021

1. Introduction

There is a large number of research studies on the importance of speaking, writing and reading, although the focus on reading skills has been increased in recent years due to its apparent importance in language teaching and learning. Reading comprehension plays a major role in language learning process. The focus of the present study was on critical thinking and inferential ability or inference with regard to reading skills.

Critical Thinking is the process of using reasoning rationale to separate what is true, and what is false. Critical thinking is a higher-order thinking skill that involves estimating opinions and is a decided, self-regulatory judgment, which ends in interpretation, analysis, evaluation, and deduction.

The analysis of the facts to form an opinion is critical thinking. The topic is complex and there are various definitions, which typically include rational, skeptical, unbiased analysis or assessment. Critical thinking is a self-driven, self-disciplined, self-corrected thinking. It requires the acceptance and careful control of rigorous standards of excellence. It involves effective communication, solving problems and a commitment for overcoming indigenous egocentrism. Critical thinking involves a disciplined, self-directed, and deliberate thinking process that encompasses 'thinking about your thinking' in an attempt to make better decisions and actions (Scheffer & Rubenfeld, 2000).

Inference is the act of coming to a consistent logical consequence without actually witnessing or having the firsthand knowledge of an event. The capacity to get to and investigate the information is very fundamental in developing knowledge. However, individuals are not generally ready to find data rapidly, assess the significance of the data, and access it (Sligo & Williams, 2002).

Among different factors that may influence reading comprehension, critical thinking and inferential ability seem to be very important. The relationship between Iranian EFL learners' critical thinking and their inferential ability in reading comprehension considering their gender and language proficiency will be investigated in this study.

Iranian EFL learners are not impressively familiar with the idea of critical thinking. Usually, they accept many ideas without making any effort to judge them in terms of truthful pieces of evidence. There is a consensus of opinion among researchers that critical thinking is important in almost every activity due to its association with some abilities such as problem solving and decision making (NourMohammadi & Zare, 2015).

This study will attempt to show the relationship between critical thinking and inferential ability in reading comprehension. Through the enhancement of these two abilities, students and teachers will be more efficient in learning and teaching. Critical thinking and inferential ability are two important abilities, which are vital in education, so a significant relationship between these two abilities can help learners acquire a better understanding of implicit meanings of what they study. Moreover, with a better understanding of critical thinking and inferential ability, teachers can design and implement appropriate reading comprehension texts to promote the students' understanding of such texts.

The current study can assist educational decision makers to recognize to what extent Iranian EFL learners have critical thinking and inferential ability in reading comprehension and / or if there is any relationship between critical thinking and inferential ability in reading comprehension. It can assist teachers to become conscious of the significance of learners' critical thinking and inferential ability and teach them how to use these abilities to be successful in different processes of learning. Norris and Phillips (1987) mention that reading is more than just saying what is on the page; it is thinking. Moreover, Beck (1989) claim "there is no reading without reasoning" (p. 677). The following research null hypotheses, based on the above questions, were posed to be tested.

1. Null hypothesis 1: There is no significant relationship between EFL learners' critical thinking and their inferential ability in reading comprehension.
2. Null hypothesis 2: There is no significant difference between Iranian EFL learners' critical thinking and their inferential ability in reading comprehension considering their gender.

3. Null hypothesis 3: There is no significant difference between Iranian EFL learners' critical thinking and their inferential ability in reading comprehension considering their language proficiency.

2. Review of Literature

Several studies have empirically explored the relationship between critical thinking and reading comprehension. For instance, Mohseni et al. (2020) examined the differential effect of metacognitive strategy training and critical thinking awareness-raising on EFL learners' reading comprehension of general reading (GR), cause and effect (CE), and argumentative (ARG) texts. The study investigated the extent to which raising Iranian EFL learners' awareness of critical thinking skills and three sets of metacognitive reading strategies including global, problem-solving, and support influence three groups of participants' reading comprehension. To this end, a purposeful sample of 54 intermediate male Iranian EFL learners attending three classes was recruited from a pool of 85 English learners and was randomly assigned to two experimental groups receiving metacognitive training (MCG) and critical thinking awareness-raising (CTG) and a control group (CG). The research data were gleaned using a researcher-made piloted 54-item reading comprehension test. The results of the paired sample t-test analyses demonstrated that both MCG and CTG made significant improvements in comprehending GR and ARG texts from the pre-test to the post-test. Moreover, both groups outperformed the CG in comprehending GR and ARG post-tests. With regard to CE comprehension, the only significant difference was observed between the MCG and the CG. The findings offer a number of pedagogical implications.

In another study, Din (2020) aim to evaluate university students' critical thinking ability as reflected in their critical reading skills. To do so, the researcher has set two research objectives to understand the university students' attitudes towards critical thinking and to know the relationship between the university students' attitudes towards critical thinking and their language proficiency with reference to their critical reading skill and their performance in critical reading test (CRT). To achieve these objectives, the researcher has used quantitative research methodology. The participants of this study consisted of 550 male and female university students of different state-run colleges of Punjab (Pakistan). Critical

thinking inventory (CTI), Watson-Glaser's Critical Thinking Appraisal (WGCTA) (2002) and critical reading test (CRT) have been used to collect data. The researcher used SPSS to analyze the collected data. The findings of the study revealed that the university students have highly positive attitudes towards critical thinking but their level of critical thinking and their ability to reflect on critical thinking in their critical reading skill do not correspond with their attitudes towards critical thinking.

Kamgar and Jadidi (2016) investigated the relationship between Iranian EFL learners' critical thinking and self-regulation and their reading comprehension ability across beginner, intermediate, and advanced levels of proficiency. To this end, 70 learners studying English Language Teaching and English Literature, who were selected through convenient sampling procedure, completed a reading placement test, a critical thinking dispositions questionnaire and a self-regulation questionnaire. The findings indicated a statistically significant correlation for those with greater critical thinking skill among advanced and intermediate learners, while there was no significant relationship between self-regulation and reading comprehension ability except for the advanced group of learners.

Nasirahmadi (2014) conducted a study to survey the relationship between Iranian language learners' critical thinking ability and their reading comprehension achievement. To do so, two B.A. English classes from the faculty of letters and humanities of Shahid Beheshti University were selected as the participants of the investigation. They were given the Persian version of Watson-Glaser Critical Thinking Appraisal (1980) and an Academic IELTS reading test. The obtained data were analyzed using Pearson-Product Moment Correlation Coefficient and doing a statistical regression analysis. The results revealed that there was a significant, positive, and moderately high correlation between the two variables of the study. The students' scores on the critical thinking test could predict their success or failure in the IELTS academic reading section.

Also, Ghabanchi and Behrooznia (2014) studied the impact of brainstorming on reading comprehension and critical thinking ability of EFL learners. This study investigated the effect of brainstorming as a pre-reading strategy on the reading comprehension ability as well as critical thinking ability of EFL learners. The study used an experimental design with

29 participants in the control group and 25 participants in the experimental one. The results of the pretest affirmed the homogeneity of the participants in two groups regarding their reading comprehension ability as well as critical thinking ability. A 45 multiple-choice reading test taken from Flash (2005) TOEFL Reading and the 'Watson-Glaser Critical Thinking Appraisal' (CTA) were implemented to assess the students' reading comprehension ability, besides their CT ability; respectively. The post-test results demonstrated that brainstorming strategies have a positive significant effect on both CT ability as well as reading comprehension ability of the participants.

Forood and Farahani (2013) did a comparative study on the performance of Iranian high and low critical thinkers on different types of reading comprehension questions. The study aimed to detect whether there was any significant difference between the achievement of high and low critical thinkers on accurate, referential, and inferential reading comprehension questions. To this end, 42 learners were selected. Afterward, the Persian version of Watson-Glaser Critical Thinking Appraisal (1980) was administered to the participants and they were then assigned to two groups of high and low critical thinkers. The findings showed that (a) there was a significant difference between the high and low critical thinkers in their performance on factual reading comprehension questions, (b) there was a significant difference between high and low critical thinkers in their performance on referential reading comprehension questions, and (c) there was a significant difference between high and low critical thinkers in their performance on inferential reading comprehension questions.

Furthermore, Basaraba et al. (2013) examined the structure of reading comprehension questions and wanted to see if literal, inferential, and evaluative comprehension truly exist. The reasons for this investigation were twofold: (a) to examine the relative difficulty of items written to assess literal, inferential, and evaluative comprehension, and (b) to compare single factor and bifactor models of reading comprehension to decide if the items were written to assess the students' literal, inferential, and evaluative comprehension abilities. Information from around 2,400 fifth grade students gathered in the fall, winter, and spring was utilized to examine these issues. The findings showed that (a) the relative difficulty of item types might

be curvilinear, with literal items being significantly less challenging than inferential and evaluative items, and (b) literal, inferential, and evaluative comprehension measurement factors clarified unique portions of difference in addition to a general reading comprehension factor.

Also some papers have been written about inferential ability and reading comprehension. Yeari (2017) studied the role of working memory in inference generation during reading comprehension: retention, (re) activation, or suppression of verbal information. This study explored the role of working memory in online activation of bridging and predictive inferences during reading comprehension. Different patterns of results were observed for high- and low-span groups only when the participants were divided based on the listening-span test. The high-span participants generated predictive inferences faster than the low-span participants. Moreover, the high-span participants generated more bridging inferences than the low-span participants, possibly due to an enhanced retention and reactivation of inference-evoking textual information.

Besides, Saadatnia et al. (2017) inspected levels of reading comprehension across text types. They explored two levels of reading comprehension, namely literal and inferential, of two text types of narration and exposition in Iranian EFL learners. One hundred eighty upper-intermediate EFL learners were given reading passages followed by both literal and inferential multiple-choice items tests. The outcomes showed that the participants consciously outperformed on the expository texts at the level of literal comprehension. Yet, regarding inferential comprehension, there was no significant difference between the two text types. The findings, from an intra-text-type perspective, literal comprehension meaningfully outweighed inferential comprehension in the expository texts, whereas no significant difference was observed between literal and inferential comprehension in the narrative texts.

In addition, Daugaard et al. (2017) analyzed the relationship between inference making, vocabulary knowledge, and the verbal working memory of 62 sixth graders (aged 12) in reading comprehension. The results showed that it may be conceivable to enhance the effect of such vocabulary instruction on reading comprehension if the interplay between vocabulary and inference making is accentuated during instruction. To conclude, inference

making was found to increase the effect of vocabulary on comprehension. This demonstrated that word meanings are significant for comprehension in the context of inference making.

Javed et al. (2016) identified reading strategies to teach literal, reorganization, and inferential comprehension questions to ESL students. This study aimed at identifying reading strategies used by teachers that help them teach literal, reorganization, and inferential comprehension questions to ESL students. The target population was the ESL teachers teaching at a secondary school level in Penang, Malaysia. A questionnaire comprising reading strategies that help ESL teachers teach literal, reorganization, and inferential comprehension questions and developed by the researchers was used to collect data. The results, based on the quantitative analysis of the data, revealed that the ESL teachers use different reading strategies for teaching literal and reorganization comprehension questions but a small repertoire of reading strategies was employed for teaching inferential comprehension questions.

Moreover, Ulu (2016) implemented a structural equation model to explain the effect of fluent reading, literal comprehension, and inferential comprehension levels of elementary school 4th grade students on success in problem solving. The sample of the research was composed of 279 4th grade students at elementary school. When the research outcomes were evaluated in terms of the effect of reading comprehension skills on problem solving skills, it was discovered that fluent reading skills do not affect problem solving skills directly, but through literal and inferential comprehension. That is, literal comprehension affects problem solving success both directly and through inferential comprehension; and that inferential comprehension skill affects problem solving success directly.

Hall (2015) studied inference instruction for struggling readers. The reason for this investigation was to provide a synthesis of the research on inference instruction interventions conducted among students who are struggling readers. The purpose of this study was to discover and compare all intervention studies that evaluated the effects of inference instruction on inferential reading comprehension. In general, the findings of this synthesis suggest that inference instruction interventions can be useful both when they target prior

knowledge activation (and teach students to integrate prior knowledge with information in the text) and when they concentrate just on integrating information in the text.

Clinton, et al. (2014) discovered gender differences in inference generation by fourth-grade students. The purpose of this study was to determine if there were gender differences among elementary school-aged students regarding the inferences they generate during reading. Fourth-grade students (130 females; 126 males) completed think-aloud tasks while reading one practice and one experimental narrative text. The findings suggested that males and females differ in their use of cognitive processes that underlie reading comprehension, particularly with respect to the likelihood of retrieval of information from episodic memory.

In addition, Pérez et al. (2014) investigated the role of working memory in inferential sentence comprehension. In this experiment, they researched whether working memory capacity plays a role in our ability to answer comprehension sentences that require text data based on these types of inferences. Participants with high and low working memory span read two narratives with four paragraphs. After each paragraph was read, they were presented with four true/false comprehension sentences. Outcomes showed that only the explicative and predictive comprehension sentences required working memory: the participants with high verbal working memory were more accurate in giving clarifications and faster in making predictions in comparison with participants with low verbal working memory span; conversely, no working memory contrasts were found in the associative comprehension sentences.

3. Method

3.1. Participants

The participants of this study were chosen from the sixth semester students of English Literature and Translation in University of Isfahan, the seventh semester students of Teaching and Translation in Islamic Azad University (Khorasgan Branch), the sixth and eighth semester of Translation in Sheikh Bahaie University, the eighth semester students of Teaching in Payame Noor University, and the seventh and eight semester students of Translation in Islamic Azad University (Najafabad Branch). The participants were 127

students. The sample consisted of 48 males and 79 females. 20 males and 34 females belonged to University of Isfahan, 10 males and 16 females belonged to Islamic Azad University (Isfahan, Khorasgan Branch), 7 males and 9 females belonged to Sheikh Bahaei University, 3 males and 7 females belonged to Payame Noor University, and 8 males and 13 females belonged to Islamic Azad University (Najafabad Branch). All of participants were at B.A. level, at age range of 20 to 26.

3.2. Instruments

In the present study, three instruments were used for the purpose of collecting the necessary data to answer the research questions. They were a quick placement test, a critical thinking questionnaire, and an inferential reading comprehension test.

Quick Placement Test

The Quick placement test (version 2) belongs to Oxford University Press and University of Cambridge Local Examinations Syndicate. This test was divided into two parts: Part One (Questions 1-40) was given to all students. Part Two (Questions 41-60) did not start unless the students were told to do so by test supervisor. The time allowed for the test was 60 minutes. The purpose of the placement test was to identify the level of the participants.

Critical Thinking Questionnaire

The Persian version of an English Critical Thinking Questionnaire (CTQ) consisted of multiple-choice items was administrated to the participants to evaluate the participants' critical thinking ability. The researchers translated the English version of the critical thinking questionnaire into Persian to make it more comprehensible to the participants. The translated questionnaire was administered to 30 students similar to the participants of this study. The reliability of the questionnaire was found to be 0.89, which was acceptable. The questionnaire was also a paper-and-pencil instrument in which there was no time limitation for the participants to answer the questions.

Inferential Reading Comprehension Test

The inferential reading comprehension test for the participants was taken from Longman Complete Course for the TOEFL Test Preparation for the Computer and Paper Test (2001)

because the items of the inferential reading comprehension test were too many, the test was divided into two passages, one of them with four passages and the other with six passages, each test had fifteen inferential multiple-choice items, and each item was followed by four options; therefore, the inferential reading comprehension test contained thirty inferential multiple-choice items. After administering the test to 30 students similar to the participants, the reliability of the inferential reading comprehension test was found to be 0.89, which was acceptable.

3.3. Procedures

The quick placement test, the critical thinking questionnaire, and the two separate passages of the inferential reading comprehension test were administered to 127 participants in this study. All the instruments were administered in four sessions. In the first session, the quick placement test was administered in 60 minutes. In the second session, the Persian critical thinking questionnaire was administered. In the third session, because the items of the inferential reading comprehension test were too many, at first, four passages of the inferential reading comprehension test were given in one session, and then the remaining passages of the inferential reading comprehension test were administered in a different session. Each session lasted about 30 minutes. Having collected the required data, the researchers started to analyze the data.

3.4. Data Analysis

The data collected were analyzed by using SPSS. To test the first null hypothesis, a Pearson product-moment correlation was run to check whether the relationship between the critical thinking and the participants' inferential ability in reading comprehension was statistically significant or not. To test the second null hypothesis, Pearson-product moment correlation and an equation using Z (Pallant, 2013, pp. 145-147) scores were used to decide if the correlations between the participants' critical thinking and their inferential ability in reading comprehension regarding gender were significantly different or not. To test the third null hypothesis, another Pearson-product moment correlation and an equation using Z scores were also used for investigating the difference between the participants' critical thinking and their inferential ability in reading comprehension regarding language proficiency.

4. Results

4.1. The First Research Hypothesis

The first null hypothesis was as follows. In order to test the first null hypothesis, the number of the participants, the means, the standard deviations (*SD*) of the inferential ability and the critical thinking tests were calculated. Table 1 shows the descriptive statistics for the scores of the inferential reading test. Table 1 shows that the distribution of the scores was normal because the significance levels (.058 and .065) are higher than $p < .05$.

H₀₁: There is no significant relationship between Iranian EFL learners' critical thinking and their inferential ability in reading comprehension.

Table 1

Descriptive Statistics for the Scores of the Inferential Reading

	N Valid	Mean	Std. Deviation
Inferential reading	127	16.62	4.21

Table 2

Tests of Normality for the Scores of the Inferential Reading Test

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	F	Sig.
Inferential Passage Scores	.094	127	.058	.980	127	.065

Table 3 shows the descriptive statistics for the scores of the critical thinking questionnaire. According to Table 3, the significance levels (.097 and .215) are higher than .05, which indicates that the distribution is normal.

Table 3

Descriptive Statistics for Scores of the Critical Thinking Questionnaire

	N	Mean	Std. Deviation
Critical thinking questionnaire	127	3.3278	.30521

Table 4

Test of Normality for the Critical Thinking Questionnaire Scores

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	Df	Sig.
Critical thinking questionnaire	.073	127	.097	.986	127	.215

The first research question put forward earlier in this study focused on the relationship between critical thinking and Iranian EFL learners' inferential ability in reading comprehension. To answer this question, first we take a look at the scatter plot in order to get a general idea of the relationship (if any) and check some assumptions. As shown in Fig 1, a positive relationship can be observed between the two variables.

As shown, looking at the graph from the bottom left corner to the top right, a cigar shape can be observed, which indicates homoscedasticity.

Next, the graph shows a few dots standing separately on their own that shows that there are just a few outliers. Finally, as can be seen, mostly higher values on the critical thinking questionnaire are associated with higher values on the inferential test suggesting a

positive linear relationship between the two. Considering these assumptions and points, we can run a Pearson's product-moment correlation analysis to further examine the relationship.

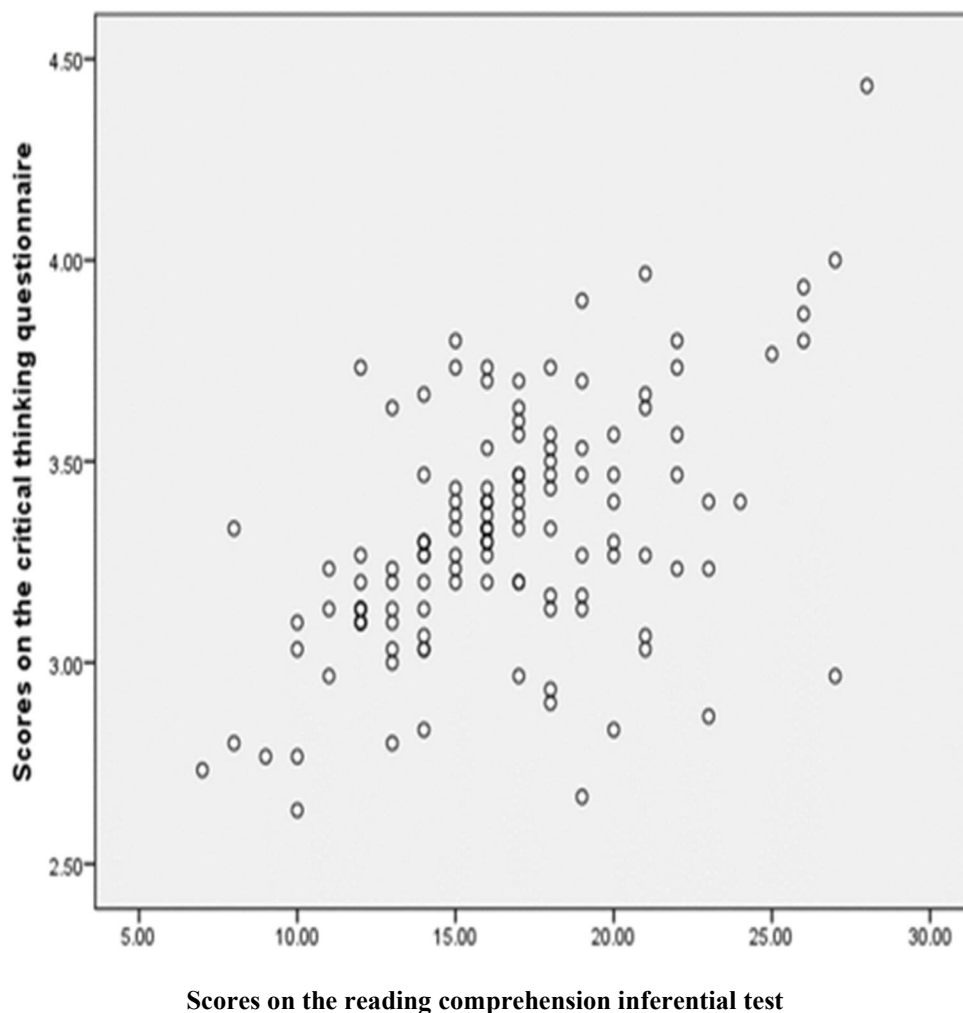


Fig 1

Scatter Plot for the Relationship between Critical Thinking and EFL Learners' Inferential Ability

Table 5 below shows the results of a Pearson's product-moment correlation procedure run on the participants' scores on the critical thinking questionnaire and the inferential reading test. As can be seen, there was a significant positive correlation between the two variables ($r = .51, p = .00$). Therefore, the first hypothesis is rejected. The amount of the

shared variance (r^2), the effect size calculated for this correlation was 27%, which is small according to Cohen's guidelines (1988, pp. 79-81).

Table 5

Correlations between the Critical Thinking and Inferential Ability of the Participants in Reading Comprehension

		Inferential ability	Critical thinking
Scores on inferential questions	Pearson Correlation	1	.516**
	Sig. (2-tailed)		.000
	N	127	127
Critical thinking	Pearson Correlation	.516**	1
	Sig. (2-tailed)	.000	
	N	127	127

4.2. The Second Research Hypothesis

The second null hypothesis was posed as follows. Table 6 and 7 show the N , M , and SD values for both male and female participants on the placement test. Table 8 shows the results of normality tests for both groups, which indicates acceptable values for normality because the significance levels (.131 and .152) are higher than .05.

H₀₂: There is no significant difference between Iranian EFL learners' critical thinking and their inferential ability in reading comprehension regarding gender.

The second research question posed in this study was to see if there was any significant difference between Iranian EFL learners' critical thinking and their inferential ability in reading comprehension regarding gender (N: Males = 48, Females = 79). To answer the second research question, the data file was split in two based on gender and then Pearson's product-moment correlation analysis was run.

Table 6

Descriptive Statistics for Males

	N Valid	Mean	Std. Deviation
Placement test	48	48.13	5.719

Table 7

Descriptive Statistics for Females

	N Valid	Mean	Std. Deviation
Placement test	79	45.09	7.465

Table 8

Tests of Normality for both Males and Females

Sex	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	Df	Sig.
Male	.207	48	.087	.865	48	.95
Female	.183	79	.131	.910	79	.152

According to Table 9, the correlation value for females ($r = .52, p = .00$) was slightly higher than the value calculated for males ($r = .50, p = .00$). Both correlation values reached the statistical significance levels. Both values indicate medium positive correlations between the participants' scores on the critical thinking questionnaire and inferential ability in reading comprehension test in terms of gender, with shared variances of 28% and 26% for females and males, respectively, which are small according to Cohen's guidelines (1988, pp. 79-81).

Next, the r values calculated for the females and males were converted into standard Z scores in order to see if the correlations for the two groups are significantly different. The data involved 48 males ($r = .50$) and 79 females ($r = .52$). Based on calculations, we can find the equivalent Z score for males ($z_1 = .56$) and females ($z_2 = .58$). Then, based on the standard

Z scores, the observed Z score was calculated so that we could compare them for significant differences. The observed Z score, in this case, was calculated to be 0.17, which means there is no statistically significant difference between the two correlation coefficients regarding gender. Therefore, the second null hypothesis is confirmed.

Table 9

Correlation between Critical Thinking and Inferential Ability in terms of Gender

Gender		inferential ability	critical thinking
Scores on inferential Questions	Pearson Correlation	1	.508
	Sig.(2-Tailed)		.000
	N	48	48
Males	Pearson Correlation	.508	1
	Critical thinking Sig.(2-Tailed)	.000	
	N	48	48
Scores on inferential Questions	Pearson Correlation	1	.526
	Sig.(2-Tailed)		.000
	N	79	79
Females	Pearson Correlation	.526	1
	Critical thinking Sig.(2-Tailed)	.000	
	N	79	79

4.3. The Third Hypothesis

The third research hypothesis was posed as follows. Table 10 shows the descriptive statistics of the placement test (N = 127). Moreover, the normality tests are presented in Table 11, which indicates a normal distribution, because the significance levels (.057 and .071) are more than .05.

H₀₃: There is no significant difference between Iranian EFL learners' critical thinking and their inferential ability in reading comprehension regarding English language proficiency.

Table 10

Descriptive Statistics for the Placement Test

	N Valid	Mean	Std. Deviation
Placement test	127	46.24	6.991

Table 11

Tests of Normality for the Placement Test Scores

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Placement Test	.197	127	.057	.893	127	.071

In addition, the descriptive statistics were also calculated. ($N_{low} = 58$, $N_{high} = 69$). Table 12 shows the values for the mean and SD. The participants were divided into high proficiency and low proficiency according to the median of the placement test. Table 13 shows the normality tests for both groups, which indicates acceptable values.

To test the third null hypothesis, the data file was split in two based on the participants' language proficiency (high and low proficiency). The participants whose scores on the placement test were above the median were considered as high proficiency and those whose scores were below the median were considered as low proficiency. Then, the

Pearson's product-moment correlation analysis was run, the results of which are shown in Table 14. As it is shown, the correlation value for the low proficiency participants ($r = .43$, $p = .00$) was lower than the value calculated for the high proficiency participants ($r = .57$, $p = .000$). Both values were statistically significant. These values indicate medium positive correlations. The shared variances calculated for these correlations were 19% and 33% for the low and high proficiency participants, respectively, which are small according to Cohen's guidelines (1988, pp. 79-81).

Table 12

Descriptive Statistics for the High and Low Proficiency Participants

Placement test	Mean	Std. Deviation
Lower Advanced	40.14	5.856
Higher advanced	51.36	1.855

Table 13

Tests of Normality for the High and Low Proficiency Students

Placement Test	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Lower-Advanced	.101	58	.200	.944	58	.159
Higher-Advanced	.145	69	.221	.916	69	.174

Next, the r values calculated above were converted into standard Z scores in order to see if the correlations for the two groups were significantly different or not. The data consisted of 58 low proficiency ($r = .43$) and 69 high proficiency ($r = .57$) participants. Based on Pallant (2013), we calculated the equivalent Z scores for the low ($z_1 = .466$) and the high proficiency ($z_2 = .655$) participants.

Then, on the basis of the standard Z scores, the observed Z scores were calculated so that we could compare them for the significant differences (if any). The observed Z score, in this case, was calculated to be -1.04, which means that there was no statistically significant difference between the two correlation coefficients regarding English language proficiency. Therefore, the third null hypothesis is also confirmed.

Table 14

Correlation between Critical Thinking and Inferential Ability in Reading Comprehension regarding Language Proficiency

Placement Test (Binned)			Scores on inferential questions	Critical thinking
Low proficiency participants	Scores on inferential questions	Pearson Correlation	1	.436**
		Sig. (2-tailed)		.001
		N	58	58
High er- proficiency participants	Critical thinking	Pearson Correlation	.436**	1
		Sig. (2-tailed)	.001	
		N	58	58
Scores on inferential questions	Scores on inferential questions	Pearson Correlation	1	.573**
		Sig. (2-tailed)		.000
		N	69	69
Scores on inferential questions	Critical thinking	Pearson Correlation	.573**	1
		Sig. (2-tailed)	.000	
		N	69	69

5. Discussion

The first research question was answered based on the relationship between critical thinking and Iranian EFL learners' inferential ability in reading comprehension. As it was indicated in the previous section, there was a medium, positive, and significant relationship between critical thinking and inferential ability in reading comprehension. The participants might have lacked experience in answering inference type of questions due to their teachers' tendency to focus on literal questions in the classroom, and the participants might have answered the questions by chance. This finding is similar to the findings of Hosseini et al. (2012), who investigated the relationship between critical thinking, reading comprehension, and reading strategies of English university students. The findings showed that there was a significant positive relationship between Iranian EFL readers' critical thinking ability and reading strategy use. Moreover, a significant positive relationship was observed between critical thinking and reading comprehension. Their results also manifested that those strategies along with critical thinking ability act as the best predictors of reading comprehension. Of course, these two studies are different to some extent, but in both studies, there was a positive relationship between critical thinking and reading comprehension.

The findings of this study also support the findings of the study done by Aloqaili (2012), who aimed at investigating the relationship between critical thinking and reading comprehension. The findings of his study revealed that there was a significant and positive relationship between reading comprehension and critical thinking. Another study with which the results of this study are in agreement is the study done by Nasirahmadi (2014). He examined the relationship between Iranian language learners' critical thinking ability and their reading comprehension achievement. The results showed that there was a positive and moderately high correlation between the two variables of the study.

When it comes to the second research question, the relationship between critical thinking and inferential ability in reading comprehension of males and females was positive and medium. It was also shown that the correlation value for the females was slightly higher than the value calculated for the males, both of which reached the statistical significance. The findings also showed that there was no statistically significant difference between the two

correlation coefficients in terms of gender. A good reason for this result might be that the participants of these studies were at the same range of age. Another reason might be that the type of instruction they received was more or less the same.

These findings are in line with the findings of Yousefi and Mohammadi (2016) who investigated the role of critical thinking skills in EFL learners' reading comprehension, and attempted to study the possible differences between males and females. The results of the present study and their study are the same

In another study, Ghorban DordiNejad and Heydari (2012) examined whether there was any relationship between the learners' critical thinking ability and their reading comprehension ability, in general, and the micro-skills of reading comprehension (i.e. inference, main idea, and specific details), in particular. Also, this study made an attempt to investigate the differences between the learners' critical thinking ability and their reading comprehension skills with regard to their gender. The results showed that there was a strong, positive correlation between the two variables with the high level of reading comprehension associated with high level of critical thinking. Moreover, it showed that there was a strong, positive correlation between critical thinking and reading comprehension micro-skills. On the other hand, the study indicated that the differences between the male and female students' micro-skills were not significant.

Regarding the third research question, there is a significant, medium, and positive correlation between critical thinking and inferential ability in terms of language proficiency. It is shown that the correlation value for the low proficiency participants was lower than the value calculated for the high proficiency participants. One reason for these findings might be that the high and low proficiency learners received the same type of instruction in terms of reading comprehension. Another reason might be that none of the two groups received any instruction in critical thinking ability. This finding is in line with the finding of Barjasteh and Vaseghi (2012) who investigated the role of critical thinking skills in EFL learners' reading comprehension performance using Bloom's taxonomy. In this study, the role of critical thinking strategies training over two language proficiency levels, high and low, was investigated. Each proficiency group was divided into a critical and non-critical group. The

outcomes demonstrated that critical thinking skills altogether influenced EFL learners' reading comprehension performance. Nonetheless, the impact of critical thinking strategy training did not differ across different language proficiency levels.

Moreover, Kamgar and Jadidi (2016) investigated the relationship between Iranian EFL learners' critical thinking and self-regulation with their reading comprehension ability. The purpose of this study was to explore the relationship between Iranian EFL learners' critical thinking and self-regulation with their reading comprehension ability across beginner, intermediate, and advanced levels of proficiency. The results indicated that there was a significant relationship between critical thinking and reading comprehension ability of the intermediate and advanced learners. Because of using a quick placement test for language proficiency, using an insufficient number of inferential reading questions, using a non-random method of sampling, and using students from specific universities and areas, the results should be generalized with caution.

6. Conclusion

The current study attempted to determine whether there is any statistically significant relationship between critical thinking and EFL learners' inferential ability in reading comprehension. It also attempted to see if there was any statistically significant difference between Iranian EFL learners' critical thinking and their inferential ability in reading comprehension regarding gender. Moreover, it attempted to see if there was any statistically significant difference between Iranian EFL learners' critical thinking and their inferential ability in reading comprehension regarding language proficiency.

The findings showed that there was a medium, positive, and significant relationship between the critical thinking and inferential ability in reading comprehension. Furthermore, the results revealed that there was no statistically significant difference between the two correlation coefficients regarding gender and language proficiency.

Therefore, it can be concluded that improvement in critical thinking and inferential ability can lead to improvement in reading comprehension. As a result, teachers should

instruct the language learners to think critically and inferentially while reading, if they intend to become proficient readers of English texts.

Conflict of interest

The author(s) certify/certifies that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in the present research paper.

References

- Aloqaili, S. (2012). The relationship between reading comprehension and critical thinking: A theoretical study. *Languages and Translation*, 24(1), 35–41. Doi: <https://doi.org/10.1016/j.jksult.2011.01.001>.
- Barjasteh, H., & Vaseghi, R. (2012). Critical thinking: A reading strategy in developing English reading comprehension performance. *Sheikhbahae EFL Journal*, 1(2), 21-34. Doi: <https://doi.org/10.22034/EFL.2012.79173>.
- Basaraba, D., Yovanoff, P., Alonzo, J., & Tindal, G. (2013). Examining the structure of reading comprehension: do literal, inferential, and evaluative comprehension truly exist?. *Reading and Writing*, 26(3), 349-379. Doi: <https://doi.org/10.1007/s11145-012-9372-9>.
- Beck, I.L. (1989). Reading and reasoning. *The Reading Teacher*, 42, 676-682. Doi: <https://doi.org/stable/20200273>.
- Clinton, V., Seipel, B., van den Broek, P., McMaster, K. L., Kendeou, P., Carlson, S. E., & Rapp, D. N. (2014). Gender differences in inference generation by fourth-grade students. *Journal of Research in Reading*, 37(4), 356-374. Doi: <https://doi.org/10.1111/j.1467-9817.2012.01531.x>.
- Daugaard, H. T., Cain, K., & Elbro, C. (2017). From words to text: inference making mediates the role of vocabulary in children's reading comprehension. *Reading and Writing*, 30(8), 1773-1788. Doi: <https://doi.org/10.1007/s11145-017-9752-2>.
- Din, M. (2020). Evaluating university students' critical thinking ability as reflected in their critical reading skill: A study at bachelor level in Pakistan. *Thinking Skills and Creativity*, 35, 100627. Doi: <https://doi.org/10.1016/j.tsc.2020.100627>.
- Forood, S. M., & Farahani, A. A. K. (2013). A comparative study between the performance of Iranian high and low critical thinkers on different types of reading comprehension questions. *Theory and Practice in Language Studies*, 3(9), 1710. Doi: <https://doi.org/10.4304/tpls.3.9.1710-1716>.

- Ghabanchi, Z., & Behrooznia, S. (2014). The impact of brainstorming on reading comprehension and critical thinking ability of EFL learners. *Procedia-Social and Behavioral Sciences*, 98, 513-521. Doi: <https://doi.org/10.1016/j.sbspro.2014.03.447>.
- Ghorbandordinejad, F., & Heydari, M. (2012). On the relationship between Iranian EFL students' critical thinking ability and their reading comprehension micro-skills. *Philippine ESL Journal*, 9, 18-22.
- Hall, C. S. (2015). Inference instruction for struggling readers: A synthesis of intervention research. *Educational Psychology Review*, 28(1), 1-22. Doi: <https://doi.org/10.1007/s10648-014-9295-x>.
- Hosseini, E., Bakhshipourkhodaei, F., Sarfallah, Sh., & Dolatabadi, H. R. (2012). Exploring the relationship between critical thinking, reading comprehension and reading strategies of English university students. *World Applied Science Journal*, 17(10), 1356-1364.
- Javed, M., Eng, L. S., Mohamed, A.R., & Ismail, S. A. M. M. (2016). Identifying Reading Strategies to Teach Literal, Reorganisation and Inferential Comprehension Questions to ESL Students. *The Journal of Asia TEFL*, 13 (3), 204-220. Doi: <https://doi.org/10.18823/asiatefl.2016.13.3.3.204>.
- Kamgar, N., & Jadidi, E. (2016). Exploring the Relationship of Iranian EFL Learners Critical Thinking and Self-regulation with their Reading Comprehension Ability. *Procedia-Social and Behavioral Sciences*, 232, 776-783. Doi: <https://doi.org/10.1016/j.sbspro.2016.10.105>.
- Mohseni, F., Seifoori, Z., & Ahangari, S. (2020). The impact of metacognitive strategy training and critical thinking awareness-raising on reading comprehension. *Cogent education*, 7(1), 1720946. Doi: <https://doi.org/10.1080/2331186X.2020.1720946>.
- Nasirahmadi, A. (2014). The relationship between Iraninan language learners' critical thinking ability and their reading comprehension achievement. *International Journal of Research Studies in language Learning*, 3 (7), 15-27. Doi: <https://doi.org/10.5861/ijrsl.2014.689>.

- Norris, S.P., & Phillips, L.M. (1987). Explanations of reading comprehension: Schema theory and critical thinking theory. *Teachers College Record*, 2, 281-306. Doi: <https://doi.org/10.1177/016146818708900205>.
- NourMohammadi, E., & Zare, Z. (2015). The relationship between critical thinking ability and listening comprehension ability of Iranian EFL learners. *International Journal of Research Studies in Language Learning*, 4(3), 47-59. Doi: <https://doi.org/10.5861/ijrsl.2014.903>.
- Pérez, A. I., Paolieri, D., Macizo, P., & Bajo, T. (2014). The role of working memory in inferential sentence comprehension. *Cognitive processing*, 15(3), 405-413. Doi: <https://doi.org/10.1007/s10339-014-0611-7>.
- Saadatnia, M., Ketabi, S., & Tavakoli, M. (2017). Levels of reading comprehension across text types: A comparison of literal and inferential comprehension of expository and narrative texts in Iranian EFL learners. *Journal of psycholinguistic research*, 46(5), 1087-1099. Doi: <https://doi.org/10.1007/s10936-017-9481-3>.
- Scheffer, B. K., & Rubenfeld, M. G. (2000). A consensus statement on critical thinking in nursing. *Journal of Nursing Education*, 39(8), 352-359. Doi: <https://doi.org/10.1046/j.1440-172x.2002.00340.x>.
- Sligo, J., & Williams, D. (2002). Critical thinking in nursing education: A literature review. *Journal of Nursing Education*, 41(2), 69-74. Doi: <https://doi.org/10.1046/j.1440-172x.2002.00340.x>.
- Ulu, M. (2016). A Structural Equation Model to Explain the Effect of Fluent Reading, Literal Comprehension and Inferential Comprehension Levels of Elementary School 4 th Grade Students on Success in Problem Solving. *Education and Science*, 41(186). Doi: <https://doi.org/10.15390/EB.2016.6303>.
- Yeari, M. (2017). The role of working memory in inference generation during reading comprehension: Retention, (re)activation, or suppression of verbal information? *Journal of Psychology and Education*, 1-12. Doi: doi.org/10.1016/j.lindif.2017.04.002.

Yousefi, S., & Mohammadi, M. (2016). Critical thinking and reading comprehension among postgraduate students: The case of gender and language proficiency level. *Journal of Language Teaching and Research*, 7(4), 802-807. Doi: doi.org/10.17507/jltr.0704.23.

The Impact of Baby College Phonics System on Iranian Children EFL Learners' Phonological Awareness, Spelling, and Willingness to Communicate

Mona Taherian¹; Masoomeh Salehi²; Mahdieh Shafipoor³

1. Master of ELT, English Department, Faculty of Humanities, Islamic Azad University, Shahr-e-Qods Branch, Tehran, Iran.
mona.taherian76@gmail.com
2. Assistant Professor, English Department, Faculty of Humanities, Islamic Azad University, Shahr-e-Qods Branch, Tehran, Iran.
mas.salehi84@gmail.com (corresponding author)
3. Assistant Professor, English Department, Faculty of Humanities, Islamic Azad University, Shahr-e-Qods Branch, Tehran, Iran.
ma.shafipoor@gmail.com

Abstract

This study is an attempt to investigate the impact of Baby College Phonics system on Iranian children EFL learners' phonological awareness, spelling, and willingness to communicate (WTC). To this end, 80 EFL learners within the age range of 8 to 12 were selected, and homogenized. The participants were then assigned to an experimental group (i.e., Baby College Phonics) and a control group. Having selected the participants and assigning them to the groups, the researchers administered a phonological awareness test, a spelling test, and a WTC questionnaire. Next, the experimental group was exposed to Baby College Phonics system, whereas the control group was instructed based on a non-phonics approach. After 15 sessions, the phonological awareness test, spelling test, and WTC questionnaire were administered again. The findings revealed that Baby College Phonics system had a significant impact on Iranian children EFL learners' phonological awareness, spelling, and WTC. The findings have some fruitful pedagogical implications for Iranian policy makers, course book writers, and material developers.

Keywords: Baby College Phonics system, Phonological awareness, Spelling, WTC

Citation: Taherian, M., Salehi, M., & Shafipoor, M. (2024). *Applications of Language Studies*, 2(2), 87-116.

Received: 2024-04-10

Revised: 2024-07-18

Accepted: 2024-07-28

Article type: Research paper

P-ISSN: 2588-6975

E-ISSN: 2588-6983

Publisher: Hazrat-e Masoumeh University Press

DOI: 10.22034/jals.2024.2026311.1026

1. Introduction

There are numerous distinct skills involved in acquisition and articulation of a foreign/second language encompassing grammar, lexicon, and pragmatics. However, phonological awareness appears to be one of the most critical competencies, not only for mature learners but also kids and children, owing to its intricate amalgamation of a series of cognitive and physiological abilities (Celce-Murcia et al., 2010).

Phonological awareness, as defined by Winter and Grawunder (2012), refers to the understanding of the various attributes of a language's sound system, such as its sounds, syllable structure, and phonotactics. This concept has been identified as a reliable indicator for predicting the development of reading, vocabulary, and spelling abilities in both first and second language acquisition. Phonological awareness skills might take into account a variety of capabilities such as identification of syllables, rhymes, phonemes, and analogical skills (Baezzat et al., 2018).

It is generally agreed that phonological awareness goes hand in hand with dictation ability (Brown, 2007). Given that one of the main problems in children with spelling problems is a failure of phonological awareness, the importance of effective teaching methods consisting of phonological awareness is more evident than ever before (Taheri & Taki, 2017).

Another equally important issue in L2 learning is learners' willingness to communicate (WTC) in second language (L2) settings. WTC, as first proposed by MacIntyre et al. (1998), refers to "volitional inclination toward actively engaging in the act of communication in a specific situation, which can vary according to interlocutor(s), topic, and conversational context, among other potential situational variables" (Kang, 2005, p. 291). It is argued that the primary goal of acquiring a second language should be to cultivate in language learners the desire to actively pursue occasions for interaction, as well as the preparedness to participate in communication.

It has frequently been observed that the process of acquiring the ability to read and write is a multifaceted endeavor, particularly for novice learners, specifically those who are young and learning English as a second language. These learners must coordinate numerous

cognitive processes in order to achieve fluency in reading and accuracy in spelling. The teaching method known as systematic phonics instruction places emphasis on the mastery of correspondences between graphemes and phonemes, and the subsequent application of this knowledge in reading and spelling words (Harris & Hodges, 1995). Phonics instruction is specifically designed for early learners in the primary grades and for children who encounter challenges in the acquisition of reading skills (The National Reading Panel, 2000, as cited by Brown, 2001).

The systematic phonics instruction, as stated by Herring (2017), leads to an enhanced growth in reading. This is due to the fact that knowledge of phonics aids in the development of word recognition. The ability to recognize words increases fluency in reading, which subsequently enhances reading comprehension. This is because children are not faced with the challenge of decoding words and can instead focus their complete attention on deriving meaning from the text. Additionally, phonics instruction is thought to enhance spelling ability, as pointed out by Pinnell et al. (1994), as reading and spelling are interconnected processes. Phonics involves the combination of sounds to read printed words, while spelling entails breaking down spoken words into sounds in order to write them. To spell a word, a child must match each sound heard in the word to a corresponding spelling. However, the impact of phonics instruction on phonological awareness and dictation has not been adequately explored.

Researchers argue that phonological awareness is a powerful indicator of early reading development and is casually related to reading ability (Cheung et al., 2001). Although the relationship between phonological awareness and written language processing is well established, the impact of phonics instruction on phonological awareness has not yet been adequately explored. Some studies show that morphological awareness is a predictor of some language components including understanding the spelling system (e.g., Fowler et al., 1995; Reed, 2012) and single word reading and reading comprehension (e.g., Qian, 2002; Tyler & Nagy, 1990). However, the role of morphological awareness in listening comprehension has not been adequately explored.

Systematic instruction in phonics has been the subject of considerable research attention of both scholars and language instructors in the context of first-language education in countries such as the United States, the United Kingdom, Australia, and New Zealand. Notably, phonics instruction is compulsory in these countries. However, there is limited knowledge regarding the impact of phonics instruction on the WTC and dictation of foreign language learners. The National Reading Panel (2000, as cited by Brown, 2001) conducted a meta-analysis of 38 studies and provided strong evidence supporting the effectiveness of phonics instruction in first-language contexts, regardless of whether it was synthetic or analytic. Yet, there is a noticeable lack of research studies on phonics instruction for foreign language or second language learners. This raises the question of whether phonics instruction is equally effective in promoting phonological awareness and dictation. To bridge the gap in the literature, the following research questions are proposed:

1. Does Baby College phonics system have any statistically significant effect on Iranian children EFL learners' phonological awareness?
2.) Does Baby College phonics system have any statistically significant effect on Iranian children EFL learners' spelling?
3. Does Baby College phonics system have any statistically significant effect on Iranian children EFL learners' WTC?

The justification for the importance of this research lies in the fact that in Iran, there is a lack of belief or limited belief in the instruction of phonics and it is necessity for the development of early literacy skills in English. This may be attributed to their strong adherence to the traditional approach of whole language instruction, which does not emphasize on phonics. In this particular study, the focus is on the use of Baby College Phonics system in an Iranian context, aiming to demonstrate how an unfamiliar approach is perceived and put into practice. Consequently, the outcomes of this study hold significance as they can offer a practical model for educators in determining effective methods of teaching English as a Foreign Language (EFL). Moreover, the utilization of Baby College Phonics system in classroom setting may provide a more contextualized understanding of the actual

teaching and learning process, enabling language teachers to decipher the impact of this approach on EFL learners' phonological awareness, dictation, and WTC.

2. Review of Literature

2.1. Phonics Instruction

Phonics is a method of teaching which focuses on ways of decoding written letters and spoken sounds. This approach encourages children to decode words by letter-sounds, instead of recognizing whole words (Kilpatrick, 2020). Stahl (2004, p. 35) claims that phonics refers to “any approach in which the teacher does or says something to help children learn how to decode words”.

Phonics is also known as the alphabetic principle or alphabetic code (Martinez, 2011). In addition, it is also a guide to an alphabet symbol of transcription so, knowing the alphabet symbols will make the translation of the spoken word into its written form and vice versa more convenient (Ming & Dukes, 2010). Phonics instruction is a way of teaching which teaches children that there is a relationship between the graphemes (letters of written language) and phonemes (the individual sounds of spoken language). The aim of this method of teaching is to make learners find out that there is a relationship between written letters and spoken sounds. It also aimed to help children how to use their phonetic knowledge in reading and writing (Lonigan, 2003, as cited in Lonigan, 2007).

Phonics instruction promotes the increase of learners' awareness allowing them to perceive the sounds in association with letters, integrating phonemic awareness. According to Doty et al. (2015, p. 503), phonics instruction teaches learners “the relationship between the letters of written language and the sounds of spoken language”. Although phonics teaching is mostly used to improve literacy skills in L1, there are examples of the use of this methodology in EFL classrooms to improve reading and pronunciation skills. Ibarrola (2007) suggests the use of phonics instruction in EFL primary school contexts to improve not only reading comprehension but also pronunciation.

According to Stahl (2004), phonics instruction is an approach in which the instructor teaches learners how to decode words. Teaching sounds-symbol correspondences directly,

having students manipulate letter sounds in written words through spelling tasks, teaching patterns of the similar spelled words or anything else that helps children learn about orthographic patterns in written form are involved in this approach.

Phonics is a way of instruction which teaches learners correspondences between graphemes in written language and phonemes in spoken language and how to use these correspondences to read and spell words (Ehri, 2003). Phonics instruction is systematic when

all the major grapheme-phoneme correspondences are taught and they are covered in a clearly defined sequence including short and long vowels as well as vowel and consonant digraphs such as oi, ea, sh, th, and blends of letter-sounds that form larger subunits in words such as onsets and rimes (Ehri, 2003, p. 3).

According to Bald (2007), the systematic teaching of phonics refers to the way of connecting the sounds of spoken English with letters or groups of letters (e.g., the phoneme /k/ can be represented by c, k, ck, ch, or q spellings) and the way of blending the sounds of letters together to produce approximate pronunciations of unknown words (the sounds /p/, /e/, /n/ can be blended to pen). Phonics is systematically taught in different ways like synthetic phonics, analytic phonics, embedded phonics, analogy phonics, onset-rime phonics, and phonics through spelling.

In line with Ibarrola's suggestion, Tomás et al. (2017) carried out a study applying phonics instruction to 25 phonemes. The study included 50 Spanish participants, aged 5 and 6 years old. One of the aims of this study was to verify if there were differences in perception and production of English sounds after phonics teaching. According to the results, phonics teaching assisted the phonological production of sounds, improving pronunciation. In addition, according to a study carried out by Matos (2018), an improvement in terms of pronunciation was achieved by young EFL learners using phonics instruction. This study focused on the contrast of sounds /θ/-/ð/ and /I/-/I:/. 13 participants aged from 10 to 12 years old. Furthermore, Martinez (2011) carried out a study about the use of explicit phonics instructions in EFL classrooms, aiming to identify if it had positive effects on literacy of 85

participants, from the first grade, attending bilingual school. The author concluded that young learners' pronunciation skills improved when reading.

2.2. Phonemic Awareness

There have been many discoveries about cognitive development and what linguistic skills children need to become successful readers. Phonological awareness has been around for many years, but connecting research to colleges and schools has been slow (Kilpatrick, 2020). In the mid to late 1900's, phonological awareness was popular in some programs. Despite that fact, many educators were not taught about phonological awareness in college and are learning the importance of it while teaching (Kilpatrick, 2020).

When teachers are effectively teaching students phonological awareness they are then reaching all levels of reading abilities and helping them learn (Hatcher, 2022). The term phonemic awareness has currently become a buzzword and many educators are beginning to understand the importance of providing developmentally appropriate instruction to teach children so they are able to build upon prior knowledge to become fluent readers (Ming & Dukes, 2010).

Different levels of phonological awareness include syllable, onset-rime, and phoneme. These skills begin to be acquired and mastered in preschool and are typically achieved by late first grade to early third grade (Kilpatrick, 2020). It is important to teach phonemic awareness skills in proper order to children making them successful in phonics and identifying words (Stotsky, 2009). Children first learn the difference between words and sentences, learn to rhyme, be aware of syllables, manipulate onset and rime, and then begin to be aware of different phonemes (sounds) to begin manipulating sounds (Moats, 2009). Phonemic awareness is one part of phonological awareness, but all of the above-mentioned skills are critical to a child's reading development (Yopp & Yopp, 2009).

Phonemic awareness "refers to the specific ability to focus on and manipulate individual sounds (phonemes) in spoken word" (Lonigan, 2007, p. 1). Being able to isolate phonemes is the first task a student will become fluent with in phonemic awareness. The student will be able to identify the beginning, middle, and end sounds of a word. Next, a

student will begin to segment phonemes in single syllable words into their individual sounds. Once the student is able to segment sounds, they will begin to blend the phonemes together to produce a word (Gillon, 2005).

The most difficult part of phonemic awareness is being able to manipulate the sounds of a word. To do this, the student will replace, take away, or add sounds to a word to produce a new word (Lonigan, 2007). A student who is able to manipulate and isolate phonemes in a word is showing the highest level of phonological awareness and is preparing to become a fluent reader (Kilpatrick, 2020).

Phonemic awareness instruction is vital for all students, because it may be difficult to identify the students who are struggling readers until they have fallen behind (Kilpatrick, 2020). Ball and Rahilly (2014) made an interesting discovery when studies showed that “phonemic awareness was more predictive of reading development than students’ IQ or socioeconomic level” (p. 28). Providing phonological awareness instruction as early as preschool using playbased methods such as nursery rhymes has been effective in developing phonemic awareness to help students understand the basic phonemes in the language (Kilpatrick, 2020). The goal of a student’s mastery of phonemic awareness is that they will be able to successfully decode words and become fluent readers.

2.3. WTC

The term Willingness to communicate (WTC) was firstly employed by McCroskey and Baer (1985) in relation to interaction in first language. It was an extension of lack of inclination to communicate by Mortensen et al. (1977) about tendency to oral performance (as cited in McCroskey & Baer, 1985). McCroskey and Baer (1985) defined L1 WTC as a steady inclination toward communication when a person feels free to communicate.

Xie (2011, p. 19) treats “WTC in L1 as a personality-based, trait-like predisposition which was relatively consistent across a variety of communication contexts and types of receivers”. To put differently, while other variables such as context may influence people’s inclination to speak, individuals show steady WTC propensity across various circumstances.

According to McCroskey and Richmond (1990), there are different variables that influence willingness to communicate in first language acquisition which are communication competence, self-efficacy, introversion, communication anxiety, and cultural diversities. In the early 1990s, McCroskey and Richmond (1990) worked on WTC in the native language speakers of various first languages in Micronesia, Sweden, Puerto Rico, and Australia. They were looking for the relationship among WTC, communication anxiety, speaking anxiety, and introversion in these four nations. The results showed that the level of speaking anxiety, WTC, introversion, and communication competence was different among the people in these countries. Moreover, it was found that the relationship among these variables in different countries was different. While US learners were more willing to express themselves, Micronesian learners were very unwilling to communicate. The learners from Sweden were found to have the most level of language skill, whilst Micronesian learners had the least level of language competence. Based on the results of this study, McCroskey and Richmond (1990) proposed that no generalization can be made without referring to the culture of the people.

MacIntyre (1994) proposes a model to forecast L1 WTC. This model assumes that self-perceived communication competence and speaking anxiety perceived by learners to have straight effect on individuals' willingness to communicate. It means a mixture of high level of communicative competence and a relative shortage of speaking anxiety lead to high a level of WTC. Personality trait was another influencing variable in his model. Xie (2011, p. 68) states that "the personality trait of introversion contributes to both communication apprehension and the perception of communicative competence, and self-esteem plays a role in reducing communication apprehension".

2.4. Empirical Studies

The phonics approach is also called a systematic and explicit phonologically- based approach that values the phonological and orthographic components of words as its focus is on teaching word recognition (Honig, 2001). Phonics has been described as both "...a system for encoding speech sounds into written symbols ... [and a way of] teaching learners the

relationships between letters and sounds and how to use this system to recognize words” (Mesmer & Griffith, 2005, pp. 366–367).

Empirical evidence from numerous intervention studies in English-speaking children (e.g., Conrad & Levy, 2011; Hines, 2009; O'Connor, 2014; Sénéchal et al., 2012) have already demonstrated the functional importance of the rime-based phonics intervention for improving children’s reading performances. In spite of this fact, very little has been known so far about the effects of such an instruction on the reading growth of ESL/EFL children since there is a lack of research in this area. Few studies have been attempted in ESL/EFL settings to explore the impact of phonics instruction on other variables such as spelling, WTC or motivation.

Agegnehu et al. (2023) aimed to assess whether a training-based intervention using an explicit rime-based phonics method improves the phonological awareness outcomes of EFL children in Ethiopia. The participants were two groups of 3rd grade children (N = 70) from two public primary schools in Hawassa, Ethiopia. A quasi-experimental research design was used in which a group of children (N = 35) were randomly assigned to the experimental group and another group (N = 35) to the control group. The experimental group was explicitly exposed to a training-based beginning reading intervention using an explicit rime-based phonics method. The control group stayed in their regular classroom with their regular beginning reading lessons with the conventional way. The phonological awareness test (PAT) was used to measure the phonological awareness outcomes in the two groups before and after the intervention. The findings showed that the experimental group achieved significant improvement on the phonological awareness in post- test and that the magnitude of the intervention’s effect was very large for all sub-subtests and the total phonological awareness. The findings also indicated that a brief and direct beginning reading instruction with an explicit rime-based phonics method helps children in a non-English language context with a first language (L1) significantly different from English to rapidly improve their phonological awareness performances. The findings further suggested that just knowledge of letters alone is not a sufficient condition for phonological awareness to develop.

Zuriyatiaslina, et al. (2018) investigated whether or not the onset-rime instruction improves phoneme blending and word reading outcomes in Malaysian EFL learners and concluded that the use of the onset-rime instruction significantly improved students' outcomes in phoneme blending and in new word reading. Similarly, Ng and Yiakoumetti (2010, as cited in Agegnehu et al., 2023) found that a systematic and explicit rime-based phonics intervention program resulted in significant improvements in word recognition outcomes of Hong Kong Chinese-speaking children.

3. Method

3.1. Participants

In order to conduct this study, 60 male and female EFL learners within the age range of 8 to 12 studying at Baby College and Home of Language in Tehran, Iran, were selected out of 80 participants. That is, they were selected from a pool of 80 participants using a two-stage sampling procedure. All the participants in the study were from Tehran province and Persian was their native language. Additionally, it was ensured that none of the participants had ever resided in a foreign country.

The participants who were included in this study were assumed to be at the elementary level of proficiency based on the evaluation made by the institute in advance. Initially, they were conveniently chosen and subsequently homogenized based on their performance on a proficiency examination, specifically the KET Test. Hence, in order to address the homogeneity of the participants, 60 individuals out of the 80 test takers, whose scores on the homogeneity test deviated by one standard deviation above and one standard deviation below the mean ($\text{Mean} \pm \text{SD}$), were selected.

Consequently, 20 test takers who obtained exceedingly high or exceedingly low scores on the test were deemed ineligible for the present study. Subsequently, the selected participants were equally divided into two groups (i.e., the experimental group receiving phonics instruction and the control group).

3.2. Instruments

Key English Test

The KET serves as evidence of an individual's ability to utilize the English language for the purpose of communicating in basic situations. The examination evaluates all four essential language skills, namely reading, speaking, writing, and listening. The reading and writing sections assess an individual's understanding of simple written information, such as signs, brochures, newspapers, and magazines, comprising a total of 56 questions. The listening section, on the other hand, consists of 25 questions and necessitates the examinee's comprehension of announcements and other oral materials delivered at a reasonably slow pace. The speaking section gauges an individual's capacity to engage in a conversation by responding to and posing simple inquiries. Notably, the speaking test is administered face to face with one or two other candidates and two examiners, which enhances the test's authenticity and reliability.

Phonological Awareness Skill Test

In this study, the Phonological Awareness Skills Test (P.A.S.T.) was administered as the pre-test and post-test in order to collect data. The test, which is a kind of achievement test, was developed by Yvette Zgonc (2000; as cited by Anthony & Francis, 2005). The appropriateness and content validity of the revised forms were ensured via expert judgement from two academic staff in ELT at two different universities in Iran. This test can be applied to evaluate language learners' phonological awareness skills at the levels of word, syllable, onset-rime, and phoneme. The P.A.S.T. consists of 10 parts. These parts consist of sentence segmentation, rhyme recognition, rhyme production, syllable blending, syllable segmentation, syllable deletion, phoneme isolation of initial sounds, phoneme isolation of final sounds, phoneme blending, phoneme segmentation, phoneme deletion of initial sounds, phoneme deletion of final sounds, phoneme deletion of first sound in consonant blend, and phoneme substitution. There are 20 items in total and each correct answer equals 1 score.

Spelling Test

A diction task was administered before and after the treatment. This test was given in order to make sure that the learners learnt letter sounds and they are able to spell them correctly at the end of level. For example, students from phonics one must be able to read and write the words which are made from letters A, S, P, T, N, I.

Words were selected from textbooks or teaching packages commonly used in Baby College. The participants were therefore supposed to have encountered these words in their lessons. All items were appropriate in difficulty level. Yeung et al. (2013) used a similar task to investigate Hong Kong kindergarteners' spelling performance. In this task, each one-syllable word was presented orally by the experimenter, together with a sentence containing the word. The children were encouraged to try their best to write the word correctly. A demonstration item was given before the actual testing. One point was given for the correct writing of each word. The internal consistency was 0.87. To avoid practice effect, two parallel forms of the same test were developed and were subsequently confirmed by two experienced language testers.

Willingness to Communicate Questionnaire

To calculate WTC of the participants, a WTC questionnaire which was designed by McCroskey and Richmond (1990), was employed. The questionnaire consisted of 12 items and the participants were completely free to choose whether they want to initiate the communication or not. The employed test was based on the Persian version of the questionnaire of willingness to communicate which was translated and validated by Ganji (2002, as cited in Tavakoli, 2013).

3.3. Procedures

Prior to the initiation of instruction, a cohort comprising a total of 80 male and female English as a Foreign Language (EFL) students was enlisted to serve as the study population. The individuals were required to adhere to specific selection criteria, which stipulated the voluntary participation of students who had never resided in a foreign country. Subsequently, the Key English Test (KET) was administered to the aforementioned 80 learners with the aim of ascertaining the absence of any significant divergence among the participants. Following the assessment of homogeneity, a sample comprising 60 learners were chosen to participate in the research endeavor. This sample consisted of individuals whose scores deviated by one standard deviation above and below the mean score. Then, the participants were assigned to an experimental group (i.e., phonics instruction) and a control group.

Having selected the participants and assigning them to the groups, the phonological awareness test, spelling test, and WTC questionnaire were administered to the participants. Next, the instruction began. The experimental group was taught through phonics instruction in order to get familiar with alphabet sounds, the written form of them and spelling of the words. The students were granted magnetic letters with the intention of engaging in play and locating the accurate spelling. Persian language served as the medium of instruction. Following this, the participants were afforded the chance to juxtapose and differentiate their own spellings in relation to those depicted on the board, subsequently rectifying any errors. The learners learnt the English letter names and how to write them. In addition, the learners were accustomed to some key terms such as letter, word, change, add, and omit.

The second step of the lesson involved the identification and classification of certain words that were compiled as the student spelled them in the word formation process. This particular phase of the lesson focused on the orthographic patterns that were shared by two or more words created during the lesson. In this stage, the students actively manipulated the index cards that were created from the lesson and proceeded to sort them based on their rhyming properties. To illustrate, the instructor selected a word from the set of cards and requested the student to locate a word that either rhymed with or resembled the selected word, such as "Can you find a word that rhymes with or looks like 'cat'?"

The students in control group were exposed to non-phonemic instructions (e.g., Discovery series) through which the words were organized alphabetically by frequency. After 15 sessions, the phonological awareness test, spelling test, and WTC questionnaire were administered again.

3.4. Data Analysis

To analyze the data by running independent-samples t-test statistical procedure, first, some assumptions were checked and the results of independent-samples t-tests are presented subsequently.

Homogeneity of Error Variances

To check the homogeneity of variances, Levene's statistic was used. Leven's statistic tests the assumption that the error variance of the dependent variable is equal across groups.

Table 1

Levene's Test of Equality of Error Variances^a

	Levene's Statistic	df1	df2	Sig.
Awareness (pre-test)	.01	1	58	.89
Awareness (post-test)	.65	1	58	.42
Spelling (pre-test)	.05	1	58	.81
Spelling (post-test)	1.30	1	58	.25
WTC (pre-test)	.05	1	58	.82
WTC (post-test)	1.48	1	58	.22

As displayed in Table 1 above, the results of Levene's test were not significant for the phonological awareness pre-test ($F = 0.01$, $\text{Sig} = 0.89$, $P > .05$) and post-test ($F = 0.65$, $\text{Sig} = 0.42$, $P > .05$), the spelling pre-test ($F = 0.05$, $\text{Sig} = 0.81$, $P > .05$) and post-test ($F = 1.30$, $\text{Sig} = 0.25$, $P > .05$), the WTC pre-test ($F = 0.05$, $\text{Sig} = 0.82$, $P > .05$) and the WTC post-test ($F = 1.48$, $\text{Sig} = 0.22$, $P > .05$). Based on these results, it can be concluded that there were not any significant differences between the variances of the groups.

Linearity of Slope of Regression Lines

This assumption is checked by drawing a scatter-plot. As Figure 1 shows, there was a linear relationship between the pre-test and the post-test scores which is an indication of the fact that the assumption of linearity of regression lines was also held.

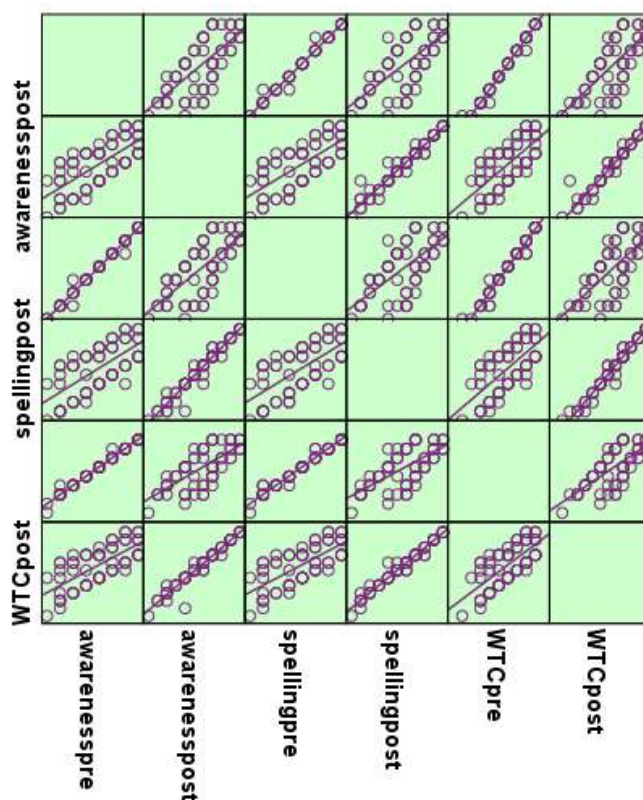


Figure 1

Linear Relationship among Regression Lines

4. Results

4.1. Testing the First Research Hypothesis

With regard to the first research hypothesis (i.e., Baby College phonics system does not have any statistically significant effect on Iranian children EFL learners' phonological awareness), the descriptive statistics showed that there was a difference between the post-test scores of the experimental group ($M=17.36$ and $SD=1.79$) and the control group ($M=14.03$ and $SD=2.09$) with regard to phonological awareness. In order to analyze whether this difference was meaningful or not, the independent-samples t-test was utilized. The results of this analysis are presented in Table 2.

Table 2

Independent Samples T-Test between the Post-tests

		Levene's Test for Equality of Variance s		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Awareness (post-test)	Equal variances assumed	.65	.42	6.63	58	.00	3.33	.50	2.32	4.33
	Equal variances not assumed			6.63	56.64	.00	3.33	.50	2.32	4.34

Based on the results presented in Table 2, it can be concluded that with the 95% confidence, there was a significant difference in the mean scores of the participants' post-tests in two groups ($t= 6.63, P<0.05$). Based on the results presented in Table 2, Baby College phonics system has a statistically significant effect on Iranian children EFL learners' phonological awareness. Thus, the first research hypothesis is rejected.

4.2. Testing the Second Research Hypothesis

With regard to the second research hypothesis (i.e., Baby College phonics system does not have any statistically significant effect on Iranian children EFL learners' spelling), the descriptive statistics showed that there was a difference between the post-test scores of the experimental group ($M=15.46$ and $SD=1.75$) and the control group ($M=11.73$ and $SD=2.11$) with regard to their spelling. In order to analyze whether this difference was meaningful or not, the independent-samples t-test was utilized. The results of this analysis are presented in Table 3.

Based on the results presented in Table 3, it can be concluded that with the 95% confidence, there was a significant difference in the mean scores of the participants' post-tests in two groups ($t= 7.43$, $P<0.05$). Based on the results presented in Table 3, Baby College phonics system has a statistically significant effect on Iranian children EFL learners' spelling. Thus, the second research hypothesis is also rejected.

4.3. Testing the Third Research Hypothesis

With regard to the third research hypothesis (i.e., Baby College phonics system does not have any statistically significant effect on Iranian children EFL learners' WTC), the descriptive statistics showed that there was a difference between the post-test scores of the experimental group ($M=27.36$ and $SD=1.79$) and the control group ($M=23.80$ and $SD=2.21$) with regard to their WTC. In order to analyze whether this difference was meaningful or not, the independent-samples t-test was utilized.

Based on the results presented in Table 4, it can be concluded that with the 95% confidence, there was a significant difference in the mean scores of the participants' post-tests in two groups ($t= 6.85$, $P<0.05$). Based on the results presented in Table 4, Baby College phonics system has a statistically significant effect on Iranian children EFL learners' WTC. Thus, the third research hypothesis is also rejected.

Table 3
Independent Samples T-Test between the Post-tests

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
Levene's Test for Equality of Variances		F	Sig.							
WTC (post-test)	Equal variances assumed	1.48	.22							
	Equal variances not assumed									
Spelling (post-test)	Equal variances assumed	1.30	.25	7.43	58	.00	3.73	.50	2.72	4.73
	Equal variances not assumed			7.43	56.9	.00	3.73	.50	2.72	4.73

Table 4

Independent Samples T-Test between the Post-tests

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		Sig.	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference			
							Lower	Upper		
TC (post-test)	qual variances assumed	.48	22	.85	8	.00	3.56	.52	.52	.60
	qual variances not assumed			.85	5.52	.00	3.56	.52	.52	.60

5. Discussion

Regarding the first research question which sought to explore if Baby College System had any statistically significant impact on Iranian children EFL learners’ phonological awareness, the analysis of the data using independent samples t-test revealed that there was a significant difference in the mean scores of the participants’ post-tests in two groups. That is, Baby College phonics system has a statistically significant effect on Iranian children EFL learners’

phonological awareness. This finding is in line with a number of research studies (e.g., Eryilmaz & Yesilyurt, 2020; Saiegh-Haddad, 2019). Furthermore, Agegnehu et al. (2023) explored whether a training-based intervention using an explicit rime-based phonics method improves the phonological awareness among EFL children in Ethiopia. He concluded that explicit rime-based phonics method assisted children in a non-English language context with a first language (L1) significantly different from English to rapidly improve their phonological awareness performances.

Regarding the second research question which inspected whether Baby College System had any statistically significant impact on Iranian children EFL learners' dictation, it can be concluded that there was a significant difference in the mean scores of the participants' post-tests in two groups. That is, Baby College phonics system has a statistically significant effect on Iranian children EFL learners' dictation. This finding corroborates with what Zuriyatiaslina et al. (2018) observed. Zuriyatiaslina et al. (2018) mentioned that the phonics approach focuses on analytical skills for breaking the code of written language. Teachers would highlight that, although the words 'big', 'pig', and 'dig' have different onsets (beginning sounds), the three words contain the same rhyme family '-ig'. Children would reflect on the shared spelling patterns across the target words. Knowledge of these patterns will help children sound out familiar words, and predict the dictation of unfamiliar words. This, in turn, would develop their dictation ability.

Finally, regarding the third research question, the findings revealed that Baby College phonics system has a statistically significant effect on Iranian children EFL learners' WTC. According to Hines (2009), vocabulary instruction can go hand-in-hand with phonics instruction. Key words that contain the target letter-sound relationships are first embedded in fun visuals that make sense to the children. Vocabulary instruction is followed by games through which children learn to identify and manipulate sounds. Once the children are able to say the whole words, the next step is to use activities to teach phonological awareness by directing the children's attention to the sound units within words (e.g., syllables, onsets, rhymes, phonemes) and encouraging the children to manipulate these sound units (e.g.,

blending and segmenting). Through this, children become more and more willing to participate in communicative interactions. Hence, the learners' WTC improves.

6. Conclusion

Based on the findings, it can be concluded that Baby College Phonics system can have a significant impact on Iranian children EFL learners' phonological awareness. It is commonly agreed that phonological awareness is seen as an important component of foreign language (L2) reading since this awareness is regarded as a pre-requirement of recognizing the relationship between a letter and its function (Saiegh-Haddad, 2019). Besides, phonological awareness can influence the writing skills of learners (Eryilmaz & Yesilyurt, 2020). As a subset of phonological awareness, phonemic awareness has occupied an important space in the relevant literature. According to Wasik (2001), phonemic awareness can be improved and assessed by teachers with the help of activities such as sound matching, sound isolation, sound blending, sound addition, substitution, or segmentation. The examples of such activities are adequately covered in Baby College Phonics system which can lay a solid foundation stone for EFL learners' phonological development.

Moreover, from the findings, it can be concluded that Baby College Phonics system can have noticeable impact on Iranian children EFL learners' spelling and WTC. Classroom instructional practices that support the development of phonics knowledge include teaching sound-letter correspondences and having students practice decoding (i.e., sound out) words (Brown, 2001). This can boost young learners' spelling ability. Furthermore, the findings may in general provide support for the observation that when children can identify words, syllables, and sounds and can make a connection between letters and sounds, they are able to write. Awareness of sounds helps them categorize orally delivered words and then write them. When children write or spell a word, they rely on the relationship between letters and sounds or phonological codes. In addition to this awareness of sounds, sound segmentation skills are essential for identifying and using unfamiliar words, for reading and writing a word, and for decoding and remembering a word. This can subsequently improve their WTC in classroom context.

The results may prove useful to policy makers as they consider the value of devoting limited instructional time and resources to teaching phonological awareness and phonics to beginning English L2 readers. At the school and classroom level, these findings may convince teachers who are uncertain about the effectiveness of phonics instructional techniques to integrate them into their L2 reading pedagogy.

Based upon the conclusions reported here, policymakers should consider providing beginning learners of English as an L2 with phonics instruction. Likewise, L2 pre-service and in-service reading teachers who are less knowledgeable about phonics instruction should seek training to teach phonics effectively to enable their learners to decode the target language better. This decoding ability is a vital first step to reading, enjoying, and ultimately learning the target language.

The format of the Baby College System facilitates the ease with which it can be implemented into the classroom. Specifically, the lessons within the Baby College System can be scripted. It contains detailed instructions about the implementation of the activities, and provides some flexibility with regards to the pacing of each lesson. Therefore, inexperienced or novice language teachers in Iran can use Baby College System with absolute confidence, being relieved that every point and activity is thoroughly explained and guided.

Baby College System can be implemented effectively in Iran educational system or any teaching programs that incorporate important literacy concepts and include activities that are enjoyable for the students. Hence, Baby College System are typically more preferred and more likely to be used by experienced and novice language teachers. Finally, Baby College System can be used with all students, both those that are demonstrating early reading skills that are considered to be at grade level as well as those that are identified as being at-risk for future reading difficulties.

Even though current research efforts have provided us with valuable insights, important work remains to be done. For example, additional research can provide us with a more fine-grained understanding of the interaction between learners' personal characteristics (e.g., age, L1 reading ability, etc.) and Baby College Phonics' instructional techniques and

approaches to which they are exposed. This study was carried out with a relatively small group of language learners. Future studies could explore the verifiability of findings with a bigger sample of study. Last but not least, future studies can inspect the role of Baby College Phonics on students' communicative apprehension and foreign language anxiety.

Conflict of interest

The author(s) certify/certifies that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in the present research paper.

References

- Agegnehu, A. Z., Bachore, M. M., & Ayele, Z. A. (2023). Effects of rime-based analogy instruction on English word recognition ability of Ethiopian children. *Journal of Language Teaching and Research*, 14(1), 20-29.
- Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science*, 14(5), 255-259.
- Baezzat, F., Moradi, M., & Motaghedifard, M. (2018). The effect of phonological awareness on the auditory memory in students with spelling problems. *Iranian Rehabilitation Journal*, 16(1), 83-90.
- Ball, M. J., & Rahilly, J. (2014). *Phonetics: The science of speech*. Routledge.
- Brown, D. (2001). *Teaching by principles: An interactive approach to language pedagogy*. Longman Pearson Education Company.
- Brown, D. (2007). *Principles of language learning and teaching*. Longman Pearson Education Company.
- Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching pronunciation hardback with audio CDs (2): A course book and reference guide*. Cambridge University Press.
- Cheung, H., Chen, H. C., Lai, C. Y., Wong, O. C., & Hills, M. (2001). The development of phonological awareness: Effects of spoken language experience and orthography. *Cognition*, 81(3), 227-241.
- Conrad, N. J., & Levy, B. A. (2011). Training letter and orthographic pattern recognition in children with slow naming speed. *Reading and Writing*, 24, 91-115.
- Doty, S. J., Hixson, M. D., Decker, D. M., Reynolds, J. L., & Drevon, D. D. (2015). Reliability and validity of advanced phonics measures. *Journal of Psychoeducational Assessment*, 33(6), 503-521.

- Ehri, L. C. (2003). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading, 18*(1), 5-21.
- Ehri, L. C. (2011). Teaching phonemic awareness and phonics in the language arts classroom: A developmental approach. In D. Lapp & D. Fisher (Eds.), *Handbook of research on teaching the English language arts* (pp. 243-249). Routledge.
- Eryilmaz, A., & Yesilyurt, Y. E. (2020). Foreign language writing as a developmental process (foundation, expansion, development, and completion): The FEDCom Model. *International Journal of Contemporary Educational Research, 7*(2), 307-334.
- Fowler, A. E., Liberman, I. Y., & Feldman, L. B. (1995). The role of phonology and orthography in morphological awareness. *Morphological Aspects of Language Processing, 4*, 157-188.
- Gillon, G. T. (2005). Facilitating phoneme awareness development in 3-and 4-year-old children with speech impairment. *ELT, 2*(2), 12-33.
- Harris, T. L., & Hodges, R. E. (1995). *The literacy dictionary: The vocabulary of reading and writing*. International Reading Association.
- Hatcher Jr, R. J. (2022). *The phonetics and phonology of Cayuga prosody* [Unpublished doctoral dissertation]. State University of New York at Buffalo.
- Herring, E. (2017). *Evaluating the effectiveness of phonics as a method for teaching reading to children and young people who experience learning difficulties and are educated in Special School settings* [Unpublished doctoral dissertation]. University of Southampton.
- Hines, S. J. (2009). The effectiveness of a color-coded, onset-rime decoding intervention with first-grade students at serious risk for reading disabilities. *Learning Disabilities Research & Practice, 24*(1), 21-32.
- Honig, B. (2001). *Teaching our children to read: The components of an effective, comprehensive reading program*. Corwin Press.

- Ibarrola, A. L. (2011). Imitating English oral texts: A useful tool to learn English Pronunciation? *Porta Linguarum: revista internacional de didáctica de las lenguas extranjeras*, 4(16), 49-63.
- Kang, S. J. (2005). Dynamic emergence of situational willingness to communicate in a second language. *System*, 33(2), 277-292.
- Kilpatrick, D. A. (2020). How the phonology of speech is foundational for instant word recognition. *Perspectives on Language and Literacy*, 46(3), 11-15.
- Lonigan, C. J. (2007). Vocabulary development and the development of phonological awareness skills in preschool children. *Vocabulary acquisition: Implications for reading comprehension. System*, 4, 15-31.
- MacIntyre, P. D. (1994). Variables underlying willingness to communicate: A causal analysis. *Communication Research Reports*, 11(2), 135-142.
- MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, 82(4), 545-562.
- Martinez, S. (2011). Dialectal variations in Spanish phonology: A literature review. *Echo*, 6(2), 1-8.
- Matos, A. E. B. D. (2018). *Influence of phonics teaching on young EFL learners' pronunciation* [Unpublished doctoral dissertation]. Universidade NOVA de Lisboa.
- McCroskey, J. C., & Baer, J. E. (1985). Willingness to communicate: The construct and its measurement. *ELT*, 3, 23-44.
- McCroskey, J. C., & Richmond, V. P. (1990). Willingness to communicate: A cognitive view. *Journal of Social Behavior and personality*, 5(2), 19-29.
- Mesmer, H. A. E., & Griffith, P. L. (2005). Everybody's selling it-But just what is explicit, systematic phonics instruction? *The Reading Teacher*, 59(4), 366-376.

- Ming, K., & Dukes, C. (2010). Gimme five: Creating a comprehensive reading lesson with all the essential elements. *Teaching Exceptional Children*, 42(3), 22-28.
- O'Connor, R. E. (2014). *Teaching word recognition: Effective strategies for students with learning difficulties*. Guilford Publications.
- Pinnell, G. S., Lyons, C. A., Deford, D. E., Bryk, A. S., & Seltzer, M. (1994). Comparing instructional models for the literacy education of high-risk first graders. *Reading Research Quarterly*, 4(3), 9-39.
- Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, 52(3), 513-536.
- Reed, D. K. (2012). *Why teach spelling? A practical approach to dictation*. Oxford University Press.
- Saiegh-Haddad, E. (2019). What is phonological awareness in L2? *Journal of Neurolinguistics*, 50, 17-27.
- Sénéchal, M. (2019). Testing a nested skills model of the relations among invented spelling, accurate spelling, and word reading, from kindergarten to grade 1. In J. Cox (Ed.), *Research in Young Children's Literacy and Language Development* (pp. 60-72). Routledge.
- Stahl, K. A. D. (2004). Proof, practice, and promise: Comprehension strategy instruction in the primary grades. *The Reading Teacher*, 57(7), 598-609.
- Stotsky, S. (2009). Licensure tests for special education teachers: How well they assess knowledge of reading instruction and mathematics. *Journal of Learning Disabilities*, 42(5), 464-474.
- Taheri, M., & Taki, S. (2017). The effect of dictogloss on listening comprehension: Focus on metacognitive strategies and gender. *International Journal of Applied Linguistics and English Literature*, 6(7), 23-29.

- Tyler, A., & Nagy, W. (1990). Use of derivational morphology during reading. *Cognition*, 36(1), 17-34.
- Wasik, B. A., & Bond, M. A. (2001). Beyond the pages of a book: Interactive book reading and language development in preschool classrooms. *Journal of educational psychology*, 93(2), 243-250.
- Winter, B., & Grawunder, S. (2012). The phonetic profile of Korean formal and informal speech registers. *Journal of Phonetics*, 40(6), 808-815.
- Xie, Q. M. (2011). *Willingness to communicate in English among secondary school students in the rural Chinese English as a foreign language (EFL) classroom* [Unpublished doctoral dissertation]. Auckland University of Technology.
- Yeung, S. S., Siegel, L. S., & Chan, C. K. (2013). Effects of a phonological awareness program on English reading and spelling among Hong Kong Chinese ESL children. *Reading and Writing*, 26, 681-704.
- Yopp, H. K., & Yopp, R. H. (2009). Phonological awareness is child's play. *Young Children*, 64(1), 12-21.
- Zuriyatiaslina, Y., Jawanees, A. H. N., Norhamimah, R., Juliana, M. N., Ngo, K. L., & Salihah, A. (2018). Improving students' blending skill using onset-rime approach. *Journal of Fundamental and Applied Sciences*, 10(2), 847-857.

Development of Intercultural Competence in Language Learners: the State of Art

Amirhossein Rastgoo 

Instructor, English Department, Islamic Azad University, Tonekabon Branch, Tonekabon, Iran.
ar.creed@yahoo.com

Abstract

This paper provides a comprehensive review of literature on the significance of intercultural competence in language learners, emphasizing its crucial role in an increasingly globalized world. Various factors that impact the development of intercultural competence, including attitudes, prior knowledge, pedagogical methods, and Byram's competence model are examined in detail. To lay the groundwork for a deeper understanding of this concept, Swain and Bachman's models of language competence are initially presented. Furthermore, the paper briefly explores some well-known perspectives and key stages in the development of intercultural competence. It is suggested that effective language programs should incorporate intercultural competence development through a multifaceted approach, while also being mindful of potential drawbacks such as the reinforcement of stereotypes or power imbalances. Additional research is warranted to investigate the influence of pedagogical strategies and other factors on the enhancement of intercultural competence. It also underscores the importance of fostering intercultural competence in language learning and underscores the necessity for educators and policymakers to prioritize this aspect when designing curricula and navigating an increasingly diverse global landscape.

Keywords: Competence, Intercultural competence, Cultural awareness, Cross-cultural communication, Critical thinking

Citation: Rastgoo, A. (2024). *Applications of Language Studies*, 2(2), 116-143.

1. Introduction

Competence is a multifaceted construct that can be defined and understood in various ways depending on the context and perspective (Orr et al., 2020). Competence refers to the ability or readiness to perform a specific task, function, or role, and it can be evaluated and measured in various ways depending on the context and purpose. In general, competence implies having the necessary capabilities, expertise, and readiness to carry out a particular activity or fulfill a specific role successfully. It goes beyond mere theoretical knowledge and encompasses the practical application and effective utilization of acquired knowledge and skills. In addition to the broad understanding of competence as the ability or readiness to perform a specific task, function, or role, there are various perspectives and models that further delve into the concept (Garcia, 2022).

Swain's Competence Model (1985) has been linked to intercultural competence development, which involves the ability to communicate effectively with individuals from different cultural backgrounds. Swain's concept of communicative competence refers to the ability of a language learner to effectively use language in real-life situations by understanding various linguistic and cultural factors and using language strategically to achieve communication goals. This involves not only language knowledge but also the ability to negotiate meaning and interact with others in meaningful ways. Studies have shown that the development of sociolinguistic and pragmatic competencies is critical for intercultural communication and that meaningful interaction with individuals from different cultural backgrounds is essential for developing intercultural competence (Bahlai et al., 2019). Swain's Competence Model also highlights the importance of meaningful interaction and the use of language for communication purposes in second language acquisition, which is also critical for intercultural competence development. Language learners who engage in authentic and meaningful communication with individuals from different cultural backgrounds are more likely to develop sociolinguistic and pragmatic competencies necessary for effective intercultural communication. In addition to language learning, there are other ways to develop intercultural competence, including exposure to diverse cultures and perspectives, cultural immersion experiences, and intercultural training programs. These

experiences can help individuals understand and appreciate different cultural perspectives, adapt to different cultural norms, and communicate effectively with individuals from diverse backgrounds (Fantini, 2009).

Another notable model which was presented by Lyle Bachman emphasizes the communicative aspect of language proficiency and posits two primary components, linguistic competence and pragmatic competence (Bachman, 1990). Linguistic competence, which refers to knowledge of grammar, vocabulary, and syntax, is important for intercultural communication as it enables individuals to use language accurately and appropriately when communicating with people from different cultures. Pragmatic competence, which refers to the ability to use language appropriately in social and cultural contexts, is also important for intercultural communication. This involves an understanding of cultural norms, values, and beliefs, as well as the ability to adapt one's communication style to fit the cultural context. There are also different statements depicting what competence is and how it is defined in scholars' viewpoints (Byram, 2020).

In psychological view, competence is the ability or capacity of an individual to perform a specific task or activity effectively, based on their skills, knowledge, and personal attributes. This view emphasizes the importance of personal qualities, such as self-confidence, motivation, and perseverance (Bandura, 1997). In educational view, competence is a measurable level of proficiency or mastery of a particular subject or skill area and focuses on the attainment of specific learning outcomes, such as knowledge, understanding, and practical application (Sparks-Langer & Colton, 1991). In cultural view, competence is the ability of an individual to function effectively in a particular cultural context and also recognizes the importance of cultural awareness, sensitivity, and adaptation (Bennett, 1993). Cultural competence refers to the ability to effectively interact with individuals from different cultural backgrounds (Betancourt et al., 2002). It involves developing knowledge, skills, and attitudes that enable individuals to communicate and collaborate across cultural boundaries and to avoid misunderstandings and conflicts (Campinha-Bacote, 2002). Having intercultural competence is also crucial in language learning as it enables successful communication and

collaboration with individuals from diverse linguistic and cultural backgrounds (Fantini, 2009).

The essentiality of intercultural competence for language learners in today's globalized world serves as a strong rationale for exploring its development. In addition, intercultural competence development can enhance language learning outcomes by improving motivation and engagement. Language learners who develop intercultural competence are even better able to connect with the language they are learning and more motivated to learn about the culture that surrounds it (Byram & Wagner, 2018). This can lead to improved language learning outcomes, as learners are more invested in their language learning journey. Reviewing intercultural competence development theories assist researchers and practitioners to gain insights into the different stages and dimensions of intercultural competence, the factors that influence its development, and the methods and techniques for measuring and developing it. This understanding can inform the design and implementation of intercultural training programs and interventions, as well as the development of assessment tools to measure intercultural competence. Reviewing the intercultural competence development theories, therewith, can help identify gaps and limitations in current research leading to new research questions and directions. It can also help researchers and practitioners to critically evaluate the existing interventions and programs and improve their effectiveness.

Traditionally, intercultural competence has been seen as the ability to understand and respect cultural differences, and to adapt one's behavior to different cultural contexts. However, a novel perspective on intercultural competence in foreign language education is to view it as a process of cultural transformation. This perspective emphasizes that learning a foreign language involves not only acquiring linguistic knowledge but also developing a new cultural identity that incorporates aspects of the target culture (Sarwari et al., 2024). According to this perspective, intercultural competence is not just about understanding and respecting cultural differences, but about actively engaging with those differences in order to transform one's own cultural identity. This involves a process of reflection and self-awareness, as students learn to critically examine their own cultural assumptions and biases,

and to question the ways in which their own culture shapes their perceptions and behaviors. Language plays a crucial role in cultural transformation, shaping identities and bridging divides. Foreign language education should focus on proficiency in navigating diverse cultural contexts to empower students in reshaping their identities (Alaei & Nosrati, 2018). Understanding the impact of context on cultural interactions, students can develop a critical understanding of intercultural competence (Ghaemi & Soltani, 2023).

2. Theoretical framework

Intercultural competence is a multifaceted concept that evolves with time and experience. Scholars have introduced various models to outline the stages of intercultural competence development. One prominent theory is the Developmental Model of Intercultural Sensitivity (DMIS) by Bennett (1986), which delineates six stages from denial to integration. Denial is the initial stage where individuals may be oblivious to cultural distinctions, perceiving their own culture as superior. Moving through stages like defense, minimization, and acceptance, individuals gradually acknowledge and value cultural differences, seeking to understand other perspectives. In adaptation stage, individuals adjust their behavior and communication to align with diverse cultural contexts. Integration, the final stage, involves a profound appreciation of cultural diversity, enabling individuals to incorporate varied perspectives into their worldview, fostering meaningful relationships with people from diverse backgrounds.

Another important theory is Communication Accommodation Theory (CAT), developed by Giles et al. (1991). CAT posits that individuals adjust their communication style to match that of their conversational partner in order to facilitate communication and build rapport. This theory emphasizes the importance of being able to adapt to different communication styles in intercultural interactions. CAT is a widely studied and applied theory in the field of intercultural communication. It suggests that individuals adjust their communication style in order to accommodate the communication style of their conversational partner. This accommodation may involve changes in language, tone, accent, nonverbal behaviors, and other aspects of communication. According to CAT, people adjust their communication style to achieve social approval, express solidarity with their conversational partner, or establish a sense of identity. The theory also suggests that

individuals may accommodate in order to reduce social distance or emphasize social distance between themselves and their conversational partner.

Stella Ting-Toomey's Face Negotiation Theory (FNT) is also an influential theory in the field (Ting-Toomey, 1988). FNT proposes that individuals from different cultures have different face-saving needs and face-giving behaviors, and that effective communication in intercultural interactions depends on being able to negotiate these differences. This theory emphasizes the importance of understanding and respecting cultural norms and expectations. The theory suggests that effective communication in intercultural interactions depends on being able to negotiate these differences. This involves understanding and respecting cultural norms and expectations, as well as being able to adapt one's communication style to meet the needs of the other person. FNT also emphasizes the importance of managing conflict in intercultural interactions. The theory suggests that conflict can arise when individuals from different cultures have conflicting face-saving needs or when one person's face-saving needs are not being respected. In order to manage the conflict effectively, it is important to understand the other person's face-saving needs and to work towards a mutually acceptable solution.

John Berry's Critical Incident Theory (CIT) proposes that intercultural competence is developed through reflective learning from critical incidents, or experiences that challenge an individual's assumptions and beliefs about other cultures (Berry, 1997). This theory emphasizes the importance of reflection and self-awareness in intercultural learning. The theory suggests that critical incidents can be positive or negative, and that they can occur in a variety of contexts, such as study abroad programs, international business ventures, or cross-cultural interactions in one's own community. The key feature of a critical incident is that it challenges an individual's assumptions and beliefs about other cultures, and forces them to reflect on their own cultural biases and values. Furthermore, CIT emphasizes the importance of reflection and self-awareness in intercultural learning. The theory suggests that individuals must reflect on their own cultural biases and assumptions in order to develop intercultural competence. This self-awareness allows individuals to recognize and challenge their own cultural biases, and to develop a more open and accepting attitude towards other cultures.

Another framework for intercultural competence development is the Cultural Intelligence (CI) model developed by Earley and Ang (2003). This model proposes four stages of intercultural competence development as follows: 1. Cognitive: the cognitive component of cultural competence refers to an individual's knowledge about different cultures and their beliefs, values, and communication styles. This knowledge can be gained through formal education, personal experiences, and exposure to diverse cultures. It involves understanding the cultural norms and expectations that shape behavior and communication, as well as recognizing cultural stereotypes and biases (Bagwe & Haskollar, 2020). This component is important for developing a foundation of understanding and respect for other cultures. 2. Physical: the physical component of cultural competence refers to an individual's ability to recognize and respond to cultural cues, such as body language and nonverbal communication. Different cultures have different norms when it comes to physical expression, such as eye contact, gestures, and personal space. Being able to recognize and interpret these cues is important for effective communication and building relationships with people from different cultures. 3. Emotional: the emotional component of cultural competence refers to an individual's ability to manage their emotions in cross-cultural interactions. This includes recognizing and adapting to cultural differences in emotional expression, such as differences in how emotions are displayed and the degree of emotional expressiveness that is considered appropriate (Bagwe & Haskollar, 2020). It also involves being aware of one's own emotional reactions and biases, and learning to manage them in a culturally appropriate way.

Cultural competence also includes a behavioral component which refers to an individual's ability to adjust their behaviors and communication styles to fit different cultural contexts (Barker, 2016). This involves being able to communicate effectively with people from different cultures, using appropriate language, tone, and nonverbal cues. It also involves understanding and respecting cultural differences in social norms, customs, and values, and adapting one's behavior accordingly. This component is important for building trust and rapport with people from different cultures, and for successfully navigating cross-cultural interactions.

The CI model emphasizes the importance of both knowledge and skills in intercultural competence development, and highlights the role of emotions in cross-cultural interactions (Earley & Ang, 2003). In addition, the evidence depict that CI model recognizes that intercultural competence is a dynamic and ongoing process that requires continuous learning and adaptation. It emphasizes the importance of self-awareness and reflection in developing intercultural competence, as well as the need for empathy and respect for other cultures. Moreover, researchers have explored the impact of Cultural Intelligence (CI) training on intercultural communication competence. Cultural intelligence points to the ability to accommodate with different cultural contexts and interact effectively with people from diverse cultural backgrounds. For example, Matsumoto et al. (2017) who conducted a study on the effectiveness of cultural intelligence training in developing intercultural communication competence found that learners who received cultural intelligence training showed significant improvement in their intercultural communication competence.

3. Review of Literature

By integrating intercultural competence development into language learning programs, educators can prepare learners to become effective communicators and collaborators in diverse cultural settings (Chen & Starosta, 2007). One example of this research is a study by Byram et al. (2018), which investigated the impact of an online intercultural learning course on the development of intercultural competence in university students. The research study showed that the language course had a beneficial impact on the intercultural competence of students, particularly in relation to their attitudes towards cultural diversity and their capacity to reflect thoughtfully on their own cultural beliefs. In general, the advancement of intercultural competence is a crucial field of investigation in language education, with significant implications for both the teaching and learning processes. By understanding how learners develop intercultural competence, educators can design more effective language programs that prepare learners for the diverse and complex global society (Barker, 2016). In addition to the study by Byram et al. (2018), there have been several other studies that have investigated the development of intercultural competence in language learners. For example, a study by Deardorff and Arasaratnam-Smith (2019) examined the impact of a study abroad

program on the development of intercultural competence in language learners. The study found that the program had a positive effect on learners' intercultural competence, particularly in terms of their knowledge of cultural differences and their ability to adapt to new cultural contexts.

Another study by Chen and Starkey (2018) explored the development of intercultural competence in Chinese language learners. The findings indicated that learners who participated in an intercultural communication course showed significant improvement in their intercultural competence, particularly in their ability to interpret and evaluate intercultural communication situations. Comprehensively, the reviewed studies suggest that intercultural competence enhancement in language learners can be facilitated through various approaches, including explicit instruction, experiential learning, and critical reflection. Moreover, the studies highlight the importance of factors such as motivation, attitudes, and prior cultural knowledge in shaping the learners' intercultural competence. As such, educators can use these findings to inform their teaching practices and design more effective language programs that promote intercultural understanding and communication. More recently, Zhang and Matsumoto (2021) found that intercultural training had a positive impact on Chinese international students' intercultural competence, especially in terms of conflict management and adaptation skills. Overall, these studies highlight the importance of intercultural training and competence in promoting successful intercultural communication, adaptation, and personal adjustment.

In Iran, Miri and Pishghadam (2018) conducted a longitudinal study that tracked the progression of intercultural competence in Iranian EFL learners over time. Their findings demonstrated that as learners engaged in language education, they exhibited noticeable improvements in various aspects of intercultural communication, such as increased empathy, better adaptation to diverse cultural contexts, and heightened awareness of cultural differences. This study underscored the transformative potential of language education in fostering intercultural competence among learners. In a related field, Khosravi and Tavakoli (2019) found a significant positive relationship between intercultural sensitivity and EFL learners' willingness to communicate. Their research highlighted how an increase in cultural

awareness and sensitivity can enhance the learners' confidence and motivation to engage in intercultural communication, ultimately leading to more effective language learning outcomes.

On the other hand, in a study conducted by Hosseini and Mohammadi (2020), the impact of study abroad experiences on intercultural competence development among Iranian university students was investigated. The research revealed that students participating in study abroad programs demonstrated notable enhancements in various aspects of intercultural skills, such as heightened cultural awareness, improved adaptability, and enhanced communication proficiency. These findings highlight the significant role of study abroad experiences in fostering intercultural competence among Iranian university students and equipping them for global citizenship.

Moreover, a study by Rahmani and Kazemi (2020) examined the influence of intercultural curricula on the enhancement of intercultural competence among Iranian high school students. The research revealed that students exposed to intercultural curricula exhibited significant improvements in various facets of intercultural skills, including increased cultural sensitivity, better cross-cultural communication abilities, and enhanced adaptability to diverse environments. These findings underscore the positive impact of integrating intercultural content into educational curricula in fostering the development of intercultural competence among Iranian high school students, preparing them for effective engagement in a globalized world.

3.1. Culture-Specific vs. Culture-General Approaches

Culture-specific approaches involve teaching about a particular culture, while culture-general approaches involve teaching about cultural dimensions and concepts that apply across cultures. According to Leask and Bridge (2013), culture-specific approaches can provide learners with deep and nuanced knowledge of a particular culture, while culture-general approaches can provide learners with a framework for understanding cultural differences and similarities. However, Deardorff (2016) argues that a combination of both approaches may be most effective in developing intercultural competence.

In recent years, scholars have continued to debate the relative merits of culture-specific and culture-general approaches to intercultural communication training. Proponents of culture-specific approaches argue that learners must develop a deep understanding of a particular culture in order to effectively communicate and work with people from that culture. For example, Hafner and Anderson (2013) argue that culture-specific knowledge is essential for building trust and rapport with intercultural partners, and that this knowledge cannot be gained effectively through culture-general training. However, other scholars have questioned the effectiveness of culture-specific approaches, particularly in today's globalized and diverse workplaces. As Deardorff (2016) notes, culture-specific knowledge can quickly become outdated or irrelevant, particularly as cultures continue to evolve and change. Furthermore, some scholars argue that focusing too heavily on one particular culture can lead to stereotypes and essentializing, which can actually hinder effective intercultural communication (Mansouri, 2013). Culture-general approaches, conversely, emphasize the development of broad cultural competencies and awareness that can be applied across a range of cultures. This can include learning about cultural dimensions, such as individualism versus collectivism, as well as developing skills such as empathy and active listening. Proponents of culture-general approaches argue that this type of training can help learners navigate cultural differences in a wide range of settings (Gudykunst, 2015). While there are ongoing debates on the effectiveness of culture-specific versus culture-general approaches, some experts have suggested that a combination of both approaches could yield the best results. For example, Kim and Park (2015) suggest that culture-specific knowledge can be integrated into broader culture-general training, allowing learners to develop a nuanced understanding of particular cultures while also building broad cultural competencies.

3.2. Short-Term vs. Long-Term Programs

Short-term programs involve brief periods of study abroad, while long-term programs involve extended periods of study abroad. According to Sasaki and Hirose (2013), short-term programs can provide learners with exposure to new cultures and perspectives, while long-term programs can provide learners with more opportunities for language and cultural immersion. However, Sasaki and Hirose (2013) argue that the effectiveness of study abroad

programs depends on a range of factors, such as the quality of the program, the level of support provided to learners, and the learners' individual characteristics and experiences. Proponents of short-term programs argue that they offer a cost-effective and accessible way for learners to gain exposure to new cultures and perspectives (Anderson et al., 2006). Short-term programs typically last between one and eight weeks and may involve language study, cultural excursions, and homestays with local families.

On the other hand, advocates for long-term study abroad programs argue that extended stays in a foreign culture can provide learners with more opportunities for language and cultural immersion, allowing them to develop a deeper understanding of the host culture (Gu & Schweisfurth, 2016). Long-term programs may last for a semester or a year, and may involve enrolling in a foreign university, conducting research, or interning with a local organization. While both short-term and long-term study abroad programs have their own advantages, research suggests that the effectiveness of these programs depends on a variety of factors. For example, Sasaki and Hirose (2013) argue that the quality of the program, the level of support provided to learners, and the learners' individual characteristics and experiences all play a role in determining the success of a study abroad program. Learners who have prior cross-cultural experience or language proficiency may be better equipped to navigate the challenges of a new culture and may benefit more from a long-term program. In addition, recent research suggests that the format of a study abroad program may be less important than the specific activities and experiences that learners engage in while abroad (Johnson, 2018). For example, participating in cultural activities, engaging with locals, and reflecting on cultural differences may be more important for developing intercultural competence than the length of the program.

3.3. Motivation and Engagement

To develop the ability to communicate effectively and appropriately with people from diverse cultural backgrounds and to understand appreciate cultural differences, learners must actively engage with individuals from different cultures. This engagement is a crucial component of achieving intercultural competence (Mirzaei & Forouzandeh, 2013). Explicit instruction may be more effective for learners who are highly motivated to learn about other cultures or have

prior cultural knowledge, as they may be more likely to engage with and benefit from the structured instruction. However, experiential learning may be more engaging and motivating for learners who prefer hands-on and immersive learning experiences. A study by Hurd and Powers (2013) showed that both explicit instruction and experiential learning can be effective for developing intercultural competence, but that learners who are highly motivated and engaged in the learning process may benefit more from experiential learning.

3.4. Context and Environment

The effectiveness of explicit instruction versus experiential learning in developing intercultural competence may vary depending on the contextual and environmental factors involved. For instance, in a classroom setting, explicit instruction may be more feasible and pragmatic, whereas in real-world intercultural contexts, experiential learning may be more appropriate. In a study conducted by Park and Kim (2016), it was revealed that the effectiveness of study abroad programs in developing intercultural competence may depend on various factors such as the program design and the degree to which learners are exposed to genuine intercultural experiences.

In recent years, there have been several studies exploring the effectiveness of explicit instruction versus experiential learning in various contexts and environments. One such study by Chun et al. (2017) examined the impact of explicit instruction versus experiential learning on second language acquisition. The researchers found that while explicit instruction was more effective in the short term, experiential learning was more effective in the long term. Another study by Cavanagh et al. (2016) explored the effectiveness of experiential learning in the context of online education. The researchers found that experiential learning activities, such as simulations and virtual labs, were effective in promoting student engagement and improving learning outcomes in online courses. In the field of business education, a study by Erden et al. (2019) compared the effectiveness of explicit instruction versus experiential learning in teaching entrepreneurship. The researchers found that while both methods were effective, experiential learning was more effective in developing entrepreneurial skills and attitudes. Furthermore, a study by Gosselin et al. (2015) explored the impact of experiential learning in the context of a professional development program for teachers. The researchers

found that experiential learning activities, such as action research and peer collaboration, were effective in improving teacher knowledge, skills, and attitudes. Overall, these studies suggest that the effectiveness of explicit instruction versus experiential learning depends on the context and environment in which the learning takes place. While explicit instruction may be more feasible and practical in certain settings, experiential learning may be more appropriate and effective in others.

3.5. Language Proficiency

According to a research study conducted by Li and Li (2016), study abroad programs can be highly advantageous for developing both language proficiency and intercultural competence, particularly for learners who possess a higher level of language proficiency. Furthermore, there has been ongoing research aimed at examining the correlation between language proficiency and intercultural competence development in relation to both explicit instruction and experiential learning. For example, a study by Kim et al. (2013) examined the effectiveness of explicit instruction in developing intercultural communication skills for learners with low levels of language proficiency. The researchers found that explicit instruction was effective in providing learners with cultural knowledge and communication strategies, which improved their ability to interact with speakers of different languages.

Another study by Wang et al. (2015) explored the effectiveness of experiential learning in developing intercultural competence of learners with high levels of language proficiency. The researchers found that experiential learning activities, such as intercultural exchange programs and service learning, were effective in providing learners with opportunities to practice and improve their language skills in a real-world intercultural setting. Moreover, a study by Deardorff and Arasaratnam-Smith (2012) highlighted the importance of integrating language learning and intercultural competence development in both explicit instruction and experiential learning. The researchers argued that language learning and intercultural competence development are interdependent processes, and that learners need to develop both to communicate and interact with speakers of different languages effectively.

3.6. Cultural Awareness and Knowledge

Cultural awareness is being conscious of cultural differences and respecting them and cultural knowledge is the understanding of different cultures, including their norms, beliefs, and practices. Both are essential for effective intercultural communication. Explicit instruction can provide learners with structured knowledge and understanding of cultural norms and practices. For example, a research study by Chen and Starosta (2013) indicated that explicit instruction can increase the learners' cultural knowledge and awareness. In contrast, experiential learning can provide learners with immersive and authentic cultural experiences that can lead to increased cultural understanding and empathy. A study by Salaberry and Kessler (2013) revealed that students who participated in a study abroad program showed greater intercultural competence, including greater cultural empathy and openness to diversity, than students who did not participate. In intercultural adaptation, the development of cultural awareness and cultural knowledge is typically seen as the first step towards intercultural competence (Bhawuk & Brislin, 2019).

Cultural awareness involves recognizing and accepting cultural differences, while cultural knowledge involves learning about other cultures and understanding how cultural values and beliefs influence behavior (Bhawuk & Brislin, 2019). Without a basic understanding of cultural differences, it is difficult for individuals to adapt their behavior and communication styles to fit different cultural contexts. Once individuals have developed cultural awareness and knowledge, they can begin to develop intercultural communication skills, such as the ability to listen actively, ask questions, and clarify misunderstandings (Kim & Kim, 2010). These skills are critical for effective cross-cultural interactions, and can help individuals navigate cultural differences and build positive relationships with people from diverse cultural backgrounds. Moreover, discourse, pragmatics, and grammar expectancy all play important roles in the development of intercultural competence, particularly in the context of intercultural communication.

Discourse refers to the way that language is used in social interactions, including the structure of conversations, the types of language used, and the conventions and norms that guide communication (Hinkel, 2017). In intercultural communication, discourse can be particularly important, as different cultures may have different expectations regarding how

communication should be structured and what types of language are appropriate in different social contexts. These differences can include the use of formal or informal language, the degree of directness or indirectness in communication, and the use of nonverbal cues. Understanding these differences in discourse can help individuals navigate intercultural communication more effectively and build stronger relationships.

Pragmatics refers to the use of language in social interactions, including the intended meaning behind language use, the cultural assumptions and norms that shape communication behaviors, and the social context in which communication takes place (Kecskes & Romero-Trillo, 2020). In intercultural communication, individuals from different cultural backgrounds may have different expectations regarding appropriate language use and may interpret language in different ways. These differences can include the use of indirect language, the appropriateness of humor in communication, and the use of nonverbal cues. Understanding these differences in pragmatics can help individuals avoid misunderstandings and build stronger relationships across cultures. Individuals who are skilled in pragmatics are better equipped to navigate intercultural communication challenges, such as misunderstandings or misinterpretations of language use. They are able to recognize and interpret nonverbal cues, such as facial expressions and gestures, and are able to use indirect language or humor in different cultural contexts appropriately. In addition, individuals with strong pragmatic skills are better able to build trust and rapport with people from different cultural backgrounds (Ivenz & Kilmova, 2022). By demonstrating an understanding and respect for cultural norms and values, they are able to establish positive relationships and avoid cultural clashes or communication breakdowns.

Grammar expectancy refers to the assumptions that individuals make about the grammar and syntax of language use, including the rules that guide language production and comprehension (Bardovi-Harlig & Mahan-Taylor, 2019). In intercultural communication, individuals from different cultural backgrounds may have different expectations regarding grammar and syntax, which can lead to misunderstandings if these expectations are not met. One example of differences in grammar expectancy across cultures is the use of indirect versus direct communication. In some cultures, such as that of the United States, direct

communication is often valued and expected. This means that individuals are expected to speak clearly and directly, and to make their intentions and meaning explicit. In contrast, in some Asian cultures, indirect communication is often valued and expected. This means that individuals may use less direct language and rely more on nonverbal cues and context to convey meaning (Jackson, 2015). Understanding these differences in grammar expectancy can help individuals communicate more effectively and avoid misunderstandings in intercultural communication. By being aware of different communication styles and norms that exist in different cultures, individuals can adapt their language use and communication style to fit the expectations of others, while also being mindful of their own cultural norms and values.

4. Discussion

In this increasingly globalized world, the acquisition of intercultural competence is crucial, and language learning plays a pivotal role in its development (Barker, 2016). There has been a recent surge in research interest in the advancement of intercultural competence in language learners, with a specific emphasis on identifying effective pedagogical methods for promoting it. This paper presents a comprehensive discussion of various factors that have an impact on the development of intercultural competence in language learners, which include their attitudes and prior cultural knowledge, as well as various pedagogical approaches such as critical reflection and experiential learning that can be utilized for promoting it. Many studies have investigated the impact of different approaches to teaching intercultural competence, including explicit instruction, experiential learning, and critical reflection (Koester & Lustig, 2015; Sarli & Phillimore, 2022). Byram's model of intercultural competence, which highlights the importance of developing skills in interpreting, relating, and discovering cultural differences, has been extensively used as a framework for teaching intercultural competence in language learning. Furthermore, several studies have underscored the significance of learners' motivation, attitudes, and prior cultural knowledge in shaping their intercultural competence development.

As highlighted in several studies, intercultural competence development is a crucial aspect of language education and study abroad experiences. Chen and Starkey (2018)

explored the enhancement of intercultural competence in Chinese language learners through an intercultural communication course. The study revealed significant improvements in learners' ability to interpret and evaluate intercultural communication situations after participating in the course. This suggests that explicit instruction, experiential learning, and critical reflection are effective approaches in fostering intercultural competence among language learners. Moreover, Miri and Pishghadam (2019) conducted a longitudinal study on Iranian EFL learners, demonstrating the positive impact of language education on intercultural competence development. The research showed that learners exhibited increased empathy, better adaptation to diverse cultural contexts, and heightened awareness of cultural differences as they progressed in their language studies. This underscores the importance of language education in not only improving language skills but also in nurturing intercultural abilities essential for effective global communication.

In a similar vein, Hosseini and Mohammadi (2020) found that Iranian university students who participated in study abroad programs experienced enhanced intercultural skills, such as improved cultural awareness and communication proficiency. These findings highlight the value of immersive experiences in foreign cultures for developing students' abilities to navigate a globalized world. Studies suggest that language education, intercultural communication courses, and study abroad programs collectively promote intercultural competence by exposing individuals to diverse cultures and fostering cross-cultural communication skills. Therefore, integrating intercultural competence development into educational curricula is crucial for preparing individuals to engage effectively in our multicultural world (Rahmani & Kazemi, 2020).

The reviewed literature also suggests that the development of intercultural competence in language learners is a complex process that requires a multifaceted approach. Effective language programs should integrate intercultural competence development into their curriculum, incorporating various pedagogical approaches to promote intercultural understanding and communication. Additionally, language teachers should cultivate a positive attitude and an open-minded approach towards cultural differences among their learners. As Chen and Starkey (2018) implied, while the importance of intercultural

competence in language learning cannot be denied, there are some potential drawbacks and challenges to its implementation. For example, the emphasis on intercultural competence may inadvertently reinforce cultural stereotypes or perpetuate power imbalances between cultures.

5. Conclusion

In conclusion, this review highlights the importance of intercultural competence development in language learners, influenced by factors like attitudes, motivation, and prior cultural knowledge. Byram's model is widely used in teaching intercultural competence. Language programs should integrate this development into their curriculum, emphasizing a multifaceted approach. Educators play a vital role in fostering positive attitudes toward cultural differences. Further research is needed to explore pedagogical approaches and other factors affecting intercultural competence. Developing intercultural competence is an ongoing process that requires dedication. Equipping language learners with these skills enables them to navigate today's diverse global society effectively.

The implications of this study are twofold. First, language teachers and program administrators can benefit from the theories reviewed in this study to design language programs that effectively promote intercultural competence development. The study can be helpful by declaring the importance of integrating intercultural competence development into language curricula, cultivating a positive attitude towards cultural differences, and utilizing various pedagogical approaches to promote intercultural understanding and communication. Language teachers and program administrators can use this information to develop comprehensive and effective language programs that develop intercultural competence in their learners. This study can also help language teachers and program administrators to better understand the challenges and barriers that learners may face in developing intercultural competence. By recognizing and addressing these challenges, teachers and administrators can create a more supportive and inclusive learning environment that fosters intercultural learning. One challenge that learners may face is cultural bias that can manifest in different ways, such as stereotyping, prejudice, or ethnocentrism. Considering the views looked at, language teachers and program administrators can help learners to recognize and overcome cultural bias by promoting self-reflection, critical thinking, and empathy. Teachers

can encourage learners to reflect on their own cultural backgrounds and biases, and to consider how these may influence their perceptions of other cultures. They can also provide opportunities for learners to engage in dialogue and exchange perspectives with people from different cultural backgrounds. Secondly, future researchers can benefit from this study by building on its suggestions and addressing some of its limitations.

Future research should focus on using robust research designs and larger sample sizes to explore the complex process of intercultural competence development in language learners. It is important to investigate the long-term effects of intercultural competence on learners' attitudes and behaviors towards cultural differences, as well as its impact on academic and professional success. Employers value intercultural competence, as it enables effective communication and collaboration across diverse cultural backgrounds, giving language learners a competitive edge in job market. Therefore, language education programs should prioritize intercultural competence development to equip learners with the skills needed to succeed in the globalized world.

Conflict of interest

The author(s) certify/certifies that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in the present research paper.

References

- Alaei, M. M., & Nosrati, F. (2018). Research into EFL teachers' intercultural communicative competence and intercultural sensitivity. *Journal of Intercultural Communication Research*, 47(2), 73–86. Doi: <https://doi.org/10.1080/17475759.2018.1424020>.
- Anderson, P. H., Lawton, L., Rexeisen, R. J., & Hubbard, A. C. (2006). Short-term study abroad and intercultural sensitivity: A pilot study. *International Journal of Intercultural Relations*, 30(4), 457-469. Doi: <https://doi.org/10.1016/j.ijintrel.2005.10.004>.
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Bagwe, T. K., & Haskollar, E. (2020). Variables impacting intercultural competence: A systematic literature review. *Journal of Intercultural Communication Research*, 49(4), 346–371. Doi: <https://doi.org/10.1080/17475759.2020.1771751>.
- Bahlai, O., Machynska, N., Matviiv-Lozynska, Y., Senkovich, O., & Voloshyn, M. (2019). Developing students' intercultural communicative competence in foreign language classroom. *Advanced Education*, 6(11), 55–59. Doi: <https://doi.org/10.20535/2410-8286.158078>.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman and Company.
- Bardovi-Harlig, K., & Mahan-Taylor, R. (2019). Developing L2 pragmatic competence. *Language Teaching*, 52(3), 365-381. Doi: <https://doi.org/10.1017/S0261444818000351>
- Barker, G. G. (2016). Cross-cultural perspectives on intercultural communication competence. *Journal of Intercultural Communication Research*, 45(1), 13–30. Doi: <https://doi.org/10.1080/17475759.2015.1104376>.
- Bennett, J. M. (1986). A developmental approach to training for intercultural sensitivity. *International Journal of Intercultural Relations*, 10(2), 179-196. Doi: [https://doi.org/10.1016/0147-1767\(86\)90005-2](https://doi.org/10.1016/0147-1767(86)90005-2).

- Bennett, J. M. (1993). Cultural marginality: Identity issues in intercultural training. In R. M. Paige (Ed.), *Education for the intercultural experience* (pp. 21-40). Intercultural Press.
- Berry, J. W. (1997). Immigration, acculturation, and adaptation. *Applied Psychology: An International Review*, 46(1), 5-34. Doi: <https://doi.org/10.1111/j.1464-0597.1997.tb01087>.
- Betancourt, J. R., Green, A. R., & Carrillo, J. E. (2002). *Cultural competence in health care: Emerging frameworks and practical approaches*. The Commonwealth Fund.
- Bhawuk, D. P. S., & Brislin, R. (2019). Cross-cultural training. In *International Encyclopedia of Intercultural Communication* (pp. 1-10). Wiley-Blackwell.
- Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Clevedon: Multilingual Matters.
- Byram, M. (2020). *Teaching and assessing intercultural communicative competence*. 2nd ed. Bristol: Multilingual Matters. Doi: <https://doi.org/10.21832/BYRAM0244>.
- Byram, M., Gribkova, B., & Starkey, H. (2018). Developing intercultural competence in university students through an online course. *Journal of Multilingual and Multicultural Development*, 39(7), 621-636. Doi: <https://doi.org/10.1080/01434632.2017.1401851>.
- Byram, M., & Wagner, M. (2018). Making a difference: Language teaching for intercultural and international dialogue. *Foreign Language Annals*, 51(1), 140–151. Doi: <https://doi.org/10.1111/flan.12319>.
- Campinha-Bacote, J. (2002). The process of cultural competence in the delivery of healthcare services: A model of care. *Journal of Transcultural Nursing*, 13(3), 181-184. Doi: <https://doi.org/10.1177/104365960201300302>.
- Cavanagh, A. J., Arbaugh, J. B., & Sridharan, S. (2016). Experiential learning activities and student engagement in online courses. *Journal of Online Learning and Teaching*, 12(3), 455-467. Doi: <https://doi.org/10.1080/01434632.2017.1401851>.

- Chen, G. M., & Starosta, W. J. (2007). Intercultural communication competence: A synthesis. In Asante M. K., Miike Y., & Yin J. (Eds.), *The Global Intercultural Communication Reader* (pp. 57–72). Routledge.
- Chen, H., & Starkey, H. (2018). Developing intercultural competence through Chinese language learning. *System*, 74(1), 98-110. Doi: <https://doi.org/10.1016/j.system.2018.04.005>.
- Chun, D. M., Kern, R. G., & Smith, B. (2017). The power of two: Deliberate practice and explicit learning in second language acquisition. *Applied Linguistics*, 38(4), 538-562. Doi: <https://doi.org/10.1093/applin/amv042>.
- Deardorff, D. K. (2016). Intercultural competence: A key to sustainable development goals. *International Journal of Intercultural Relations*, 55(2), 106-110. Doi: <https://doi.org/10.1016/j.ijintrel.2016.09.001>.
- Deardorff, D. K., & Arasaratnam-Smith, L. A. (2012). Intercultural competence: Integrating language and cultural competence. *Language Teaching*, 45(2), 133-144. Doi: <https://doi.org/10.1017/S0261444811000537>.
- Deardorff, D. K., & Arasaratnam-Smith, L. A. (2019). The impact of study abroad on the development of intercultural competence. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 31(1), 1-18. Doi: <https://doi.org/10.36366/frontiers.v31i1.409>.
- Elola, I., & Oskoz, A. (2015). Technology-enhanced language learning for intercultural competence development: A review of empirical studies. *Computer Assisted Language Learning*, 28(3), 247-265. Doi: <https://doi.org/10.1080/09588221.2014.991475>.
- Earley, P. C., & Ang, S. (2003). *Cultural intelligence: Individual interactions across cultures*. Stanford University Press.
- Erden, Z., Yıldırım-Öktem, Ö., & Gürbüz, R. (2019). Experiential learning vs. explicit instruction in entrepreneurship education: A comparison study. *Journal of*

Entrepreneurship Education, 22(3), 1-15. Doi: <https://doi.org/10.1177/1096348019857235>.

Fantini, A. E. (2009). Assessing intercultural competence. In D. K. Deardorff (Ed.), *The Sage handbook of intercultural competence* (pp. 456–476). Thousand Oaks, CA: Sage.

Garcia, J. V. (2022). Integration of intercultural communicative competence: A case of English language teachers in higher education. *English as a Foreign Language International Journal*, 26(1), 30–62. Doi: <https://doi.org/10.56498/163212022>.

Ghaemi, H., & Soltani, S. (2023). Examining intercultural communicative competence in collectivist and individualistic EFL teachers. *International Journal of Clinical and Medical Education Research*, 2(3), 61-67. Doi: <https://doi.org/10.33140/IJCMER>.

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1-68). Cambridge University Press.

Gosselin, D., Derosier, J., & Duval, D. (2015). The effectiveness of an experiential professional development program for teachers. *Journal of Research in Innovative Teaching & Learning*, 8(2), 181-203. Doi: <https://doi.org/10.1108/JRIT-07-2014-0027>.

Gudykunst, W. B. (2015). *Bridging differences: Effective intergroup communication* (4th ed.). SAGE Publications.

Gu, Q., & Schweisfurth, M. (2016). Transcending the national context: Revisiting debates on the role of international education. *Compare: A Journal of Comparative and International Education*, 46(4), 636-656. Doi: <https://doi.org/10.1080/03057925.2015.1122694>.

Hafner, C. A., & Anderson, S. S. (2013). Teaching culture and intercultural competence in a business communication course. *Business and Professional Communication Quarterly*, 76(4), 440-453. Doi: <https://doi.org/10.1177/2329490613508138>.

- Hinkel, E. (2017). *Handbook of research in second language teaching and learning*. Routledge.
- Hosseini, S., & Mohammadi, Z. (2020). The Impact of Study Abroad Experiences on Intercultural Competence Development Among Iranian University Students. *Journal of Intercultural Communication Research*, 49(4), 327-345. Doi: <https://doi.org/10.1080/17475759.2020.1802693>.
- Hurd, S., & Powers, S. M. (2013). Intercultural competence in a global world. *Journal of Studies in International Education*, 17(3), 205-215. Doi: <https://doi.org/10.1177/1028315312453927>.
- Ivenz, P., & Klimova, B. (2022). A review study of activities used in the development of intercultural communication competence in foreign language classes. *International Journal of Society, Culture & Language*, 10(2), 137-150. Doi: <http://doi.org/10.22034/ijscsl.2022.556666.2681>.
- Jackson J. (2015). Becoming interculturally competent: Theory to practice in international education. *International Journal of Intercultural Relations*, 48(2), 91–107. Doi: <https://doi.org/10.1016/j.ijintrel.2015.03.012>.
- Johnson, R. D. (2018). Critical intercultural competencies and study abroad: A review and call for research. *Journal of Studies in International Education*, 22(5), 433-450. Doi: <https://doi.org/10.1177/1028315318777005>.
- Kecskes, I., & Romero-Trillo, J. (2020). Intercultural Pragmatics. In *The Routledge Handbook of Language and Intercultural Communication* (pp. 94-110). Routledge.
- Khosravi, P., & Tavakoli, M. (2019). The Impact of Intercultural Sensitivity on EFL Learners' Willingness to Communicate. *Iranian Journal of Language Teaching Research*, 7(1), 1-18. Doi: <https://doi.org/10.30466/ijltr.2019.80237.0>.
- Kim, Y. Y., & Kim, K. S. (2010). *Communication and cross-cultural adaptation: An integrative theory*. Routledge.

- Kim, Y. Y., & Park, G. K. (2015). Intercultural communication training: A review and theoretical investigation. *International Journal of Intercultural Relations*, 48(2), 1-14. Doi: <https://doi.org/10.1016/j.ijintrel.2015.03.002>.
- Kim, Y. Y., Park, Y., & Rubdy, R. (2013). Low-level learners and intercultural communication: An empirical study of explicit instruction. *Language and Intercultural Communication*, 13(1), 64-81. Doi: <https://doi.org/10.1080/14708477.2012.758982>.
- Koester J., Lustig M. W. (2015). Intercultural communication competence: Theory, measurement, and application. *International Journal of Intercultural Relations*, 48(1), 20–21. Doi: <https://doi.org/10.1016/j.ijintrel.2015.03.006>.
- Li, L., & Li, Y. (2016). The impact of study abroad on the development of students' intercultural competence and foreign language proficiency. *Journal of Multilingual and Multicultural Development*, 37(5), 467-483. Doi: <https://doi.org/10.1080/01434632.2015.1071821>.
- Mansouri, F. (2013). Beyond culturalism: Addressing the social determinants of identity, belonging and exclusion. *Journal of Intercultural Studies*, 34(6), 583-600. Doi: <https://doi.org/10.1080/07256868.2013.836456>.
- Matsumoto, D., Yoo, S. H., & Nakagawa, S. (2017). Individual and situational predictors of cultural intelligence development: A longitudinal study. *International Journal of Intercultural Relations*, 60(2), 60-68. Doi: <https://doi.org/10.1016/j.ijintrel.2017.08.002>.
- Miri, M., & Pishghadam, R. (2018). A Longitudinal Study of the Development of Intercultural Competence in Iranian EFL Learners. *Journal of Language and Intercultural Communication*, 18(4), 476-492. Doi: <https://doi.org/10.1080/14708477.2018.1499976>.
- Mirzaei, A., & Forouzandeh, F. (2013). Relationship between intercultural communicative competence and L2-Learning motivation of Iranian EFL Learners. *Journal of*

- Intercultural Communication Research*, 42(3), 300–318. Doi: <https://doi.org/10.1080/17475759.2013.816867>.
- Orr, J. A., Sackett, P. R., & DuBois, C. L. (2020). Competence. In I. B. Weiner, R. J. Contrada, & J. F. Karoly (Eds.), *Handbook of psychology* (2nd ed., Vol. 1, pp. 257-274). Wiley.
- Park, J. Y., & Kim, Y. K. (2016). Study abroad programs and intercultural competence development: A comparative analysis of program types and duration. *Journal of Studies in International Education*, 20(2), 165-185. Doi: <https://doi.org/10.1177/1028315315626901>.
- Rahmani, A., & Kazemi, F. (2020). The impact of intercultural curricula on the development of intercultural competence among Iranian high school students. *Journal of Intercultural Education*, 37(4), 456-468. Doi: <https://doi.org/10.1080/14675986.2020.1717318>.
- Salaberry, M. R., & Kessler, G. (2013). Developing intercultural competence through education abroad: The case of US learners of Spanish. *Foreign Language Annals*, 46(1), 55-71. Doi: <https://doi.org/10.1111/flan.12019>.
- Sarli A., & Phillimore, J. (2022). The intercultural competence of second-generation individuals: Knowledge gaps and steps forward. *International Journal of Intercultural Relations*, 88(1), 11–21. Doi: <https://doi.org/10.1016/j.ijintrel.2022.03.004>.
- Sarwari, A. Q., Adnan, H. M., Rahamad, M. S., & Abdul Wahab, M. N. (2024). The Requirements and Importance of Intercultural Communication Competence in the 21st Century. *Sage Open*, 14(2). Doi: <https://doi.org/10.1177/21582440241243119>.
- Sasaki, M., & Hirose, T. (2013). The effectiveness of study abroad programs in Japan: A comparative study of short-term and long-term programs. *The Journal of Studies in International Education*, 17(1), 57-71. Doi: <https://doi.org/10.1177/1028315311432099>.

- Sparks-Langer, G. M., & Colton, A. B. (1991). Synthesis of research on teachers' pedagogical thoughts, judgments, decisions, and behavior. *Educational Leadership*, 48(6), 56-64. Doi: <https://doi.org/N/A>.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. *Input in Second Language Acquisition*, 2(1), 235-253. Doi: <https://doi.org/N/A>.
- Ting-Toomey, S. (1988). Intercultural conflict styles: A face-negotiation theory. In Y. Y. Kim & W. B. Gudykunst (Eds.), *Theories in intercultural communication* (pp. 213-235). Sage.
- Wang, Y., Lin, L., & Li, W. (2015). Experiential learning for enhancing intercultural competence and language proficiency of international students. *Journal of Experiential Education*, 38(4), 350-366. Doi: <https://doi.org/10.1177/1053825915597024>.
- Zhang, Y., & Matsumoto, D. (2021). The effect of intercultural training on intercultural competence: A longitudinal study of Chinese international students. *Journal of Cross-Cultural Psychology*, 52(2), 159-174. Doi: <https://doi.org/10.1177/0022022120974307>.

Demotivational Factors Affecting Iranian EFL Learners' Oral Performance: the Case of High Schools

Hassan Bashirnezhad¹ Zahra Vafaenezhad

1. Assistant Professor, Farhangian University, Amol, Iran.
h.bashirnezhad@cfu.ac.ir (corresponding author)

2. Master of English Language Teaching, Islamic Azad University, Sari Branch, Sari, Iran.
zahravafaenezhad@gmail.com

Abstract

The study of motivation is concerned with what makes people interested in learning a second language and what keeps them motivated to do so. In addition to factors which increase students' motivation, there are factors which reduce their motivation, that is, demotivation factors. The present study aimed at identifying the ideas of Iranian English as a Foreign Language (EFL) teachers and learners about demotivating factors affecting the learners' oral performance. 60 EFL teachers and 100 EFL learners from high schools in Sari, Iran were selected by convenience sampling. A demotivation questionnaire was used including 20 items in a Likert-scale format ranging from strongly disagree (1) to strongly agree (5). The results indicate that according to the language teachers' and learners' opinions, 'teachers' not being equipped with the required skills for teaching', "students' lack of mental health and good memory", "students' not receiving sufficient timely incentives", "teachers' lack of sufficient knowledge about the subject", and "students' lack of enough self-confidence" are the most important demotivating factors among the others.

Keywords: Attitudes, Demotivation factors, Oral performance, Iranian EFL learners, Iranian EFL teachers

Citation: Bashirnezhad, H., & Vafaenezhad, Z. (2024). *Applications of Language Studies*, 2(2), 144-168.

1. Introduction

There are diverse factors that play an important role in the process of language learning and teaching. Motivation is a major one and is usually defined as an internal state that arouses, directs, and maintains behavior. We all know how it feels to be motivated, and to move energetically toward a goal. We also know that it is something like working hard, even if we are not fascinated by the task (Woolfolk et al., 2003).

The study of motivation is concerned with what makes someone interested in learning a second language and what keeps him/her motivated. However, motivation to learn a second language (L2) is a complex construct, considering that language is always socially and culturally bound and hence quite different from other school subjects. Students who have higher levels of motivation are more successful and efficient in their learning process (Ely, 1986). Many researchers have investigated how students can be motivated (e.g., Crookes & Schmidt, 1991; Dornyei, 2001a; Oxford, 1996). Among these researchers, Dornyei, in particular, has done extensive research on practical aspects of motivation such as the question of how teachers can help to improve the learners' motivation in classroom (e.g., Dornyei, 2001b).

In addition to factors which increase the students' motivation, there are factors which more likely reduce motivation rather than reinforcing it, that is, demotivation factors. Dornyei (2001a) has defined demotivation as "specific external forces that reduce or diminish the motivation basis of a behavior intention or an ongoing action" (p. 143). Therefore, a demotivated learner is someone who was once motivated, but has lost his or her interest for some reasons. In the same vein, we can speak of demotives, which are negative counterparts of motives.

Sometimes it is said that demotivation is the same as amotivation, while some scholars like Dornyei et al. (1985) believe that amotivation is "the relative absence of motivation that is not caused by a lack of initial interest but rather by individual feelings of incompetence and helplessness when faced with the activity". Accordingly, "amotivation" events are those that occur within the person and signify his or her inability to master some situation or events" (Dornyei, et al., 1985, p. 110).

One of the skills that learners might feel unable to develop is speaking, especially in contexts like Iran, where speaking practice does not happen much. With regard to the importance of speaking, there are three major factors encouraging learners to speak in classrooms. First, speaking activities prepare practice opportunities-chances to practice real life speaking in the classroom. Secondly, speaking tasks in which students tend to use any or all of the language they know provide feedback for both teacher and students. Finally, the more students have opportunities to activate the diverse elements of language they have kept in their brains, the more automatic their use of these elements become. As a result, students gradually become independent language users. This means that they will afford to use words and phrases fluently without very much conscious thought (Harmer, 2008). However, some factors may affect the speaking practice negatively and they may decrease the learners' motivation to speak in the classroom. This is a major concern in English Language Teaching (ELT) because motivation plays an important role in learning process and the related literature shows that those students who have higher levels of motivation are more successful and efficient in their learning process (e.g., Ely, 1986). It is obvious that motivation and demotivation as well as speaking are very important matters in the process of language learning. In this way, language teachers can become aware of factors which may encourage language learners to speak and also factors which may impede their speaking progress.

Despite the importance of demotivating factors and severity of the problem of speaking instruction, few insightful studies have been carried out on this issue in the context of Iran. Hence, this study makes a deliberate effort to partly fill the existing gap. It is trying to explore the ideas of Iranian EFL language teachers and learners about the demotivating factors affecting the oral performance of Iranian EFL learners.

2. Review of Literature

2.1. Motivation and Demotivation

Motivation is an important factor in learning a second/foreign language (Gardner, 1985b; Scarcella & Oxford, 1992). Research shows that motivation has a positive effect on language learning. However, there is another side to motivation that probably every learner has encountered at some point: temporary loss of motivation. In contrast with factors that have a

positive effect on motivation, there are also some factors with a negative effect. These factors are called demotivating factors since they may negatively influence the learner's interest in language learning and development of the four skills, namely listening, speaking, reading, and writing.

Several definitions and theories have been presented with regard to motivation so far. These theories try to explain human thought and behavior. There are unlimited reasons why one can be motivated towards an aim, which vary from one individual to another, and there are also different factors which can reduce or increase motivation. Therefore, it is unsurprising that developing a decisive definition of motivation has been problematic. This, in itself, is a daring task and so we should not expect a complete and full explanation of this complicated concept (Dörnyei, 1999, 2005; MacIntyre, 2002). Pennington (1995) believes that this problem is due to the fact that motivation cannot be directly observed and that the properties of each individual's motivational behavior are different and vary under different circumstances.

Whilst acknowledging that scholars differ in what they believe constitutes motivation, Pintrich and Schunk (2002) state that, motivation, in its complexity, "has been conceptualized in varied ways including inner forces, enduring traits, behavioral responses to stimuli, and sets of beliefs and affects" (p. 5). This definition presents us with a category of variables. An added difficulty is that diverse motives, at times, positively interact, whilst at other times they are in conflict. Moreover, in many studies, according to Schmidt et al. (1996), it is not clear whether successful learning is the result or the cause of motivation. Dodick (1996) believes that a more adequate explanation of motivation has yet to be discovered and, indeed, no single definition of motivation has been agreed upon at present (Pintrich & Schunk, 2002).

2.2. Studies on Motivation and Demotivation

In 1985, Gardner defined motivation as a multi-faceted construct which combines effort, desire, and positive attitudes toward language learning if the goal of learning L2 is to be achieved. This definition of motivation generated several problems because its main concern was integrative orientation which implied a situation where the language learners were in contact with the target language community. This is not often the case and learners have

almost no contact with speakers of the target language (Dörnyei, 1990; Ryan, 2005). Also, Gardner's definition does not take into consideration the fact that motivational levels fluctuate over time. The idea of time was included in Crookes and Schmidt's (1991) expanded definition of motivation. For them, language learning motivation features both internal and external aspects. The internal aspects include an interest in L2, relevance of learning the L2 to personal needs, expectancy of success or failure, and rewards. The external aspects consist of the decision to engage in language learning process, persistence over time, and maintaining the activity at high levels.

Dörnyei (1998) presents a more dynamic definition of motivation to learn, emphasizing the fact that motivation needs to be supported over time. For Dörnyei, "L2 motivation provides the primary impetus to initiate the learning behavior and later the driving force to sustain the long and often tedious learning process; that is, all the other factors involved in L2 acquisition presuppose motivation to some extent" (Dörnyei, 1998, p. 1).

Both Chambers (2001) and Dörnyei (2001) emphasize the multifaceted nature of motivation. They see motivation as changing over time and being influenced by the surrounding context. They recognize various factors which affect human behavior and which are, therefore, related to motivation. Both authors draw up Dörnyei and Ottó's (1998) definition of motivation which takes all of these factors into consideration: "the dynamically changing cumulative arousal in a person that initiates, directs, coordinates, amplifies, terminates, and evaluates the cognitive and motor processes whereby initial wishes and desires are selected, prioritized, operationalized and (successfully or unsuccessfully) acted out" (p. 64).

According to MacIntyre (2002), any explanation of language learning motivation requires one to answer three questions:

- (a) Why is the performance directed toward a specific goal?
- (b) What establishes the amount of intensity or effort devoted to achieving the goal?
- (c) Why do people who have the same learning situations differ in their motivational levels?

Regarding the last question, MacIntyre (2002) focuses on the fact that individual differences are of extreme importance when it comes to studying motivation, and he acknowledges several authors, such as Crookes, Schmidt, Dörnyei, Oxford and Sherain, who actually promote this in their definition of motivation. This same attention to individual differences is also given by Masgoret and Gardner (2003). Recent approaches to L2 motivation have revolved around the notions of possible self-motivation conceptualized at the group level, motivational self-regulation, teacher-controlled motivational strategies, task motivation, Ushioda's model and the role of motivational thinking.

As English becomes more global, the impact of the culture or culture associated with the English language diminishes. Rather than identifying with any particular culture, one identifies with a 'future' self which will be proficient in English (Ryan, 2005). This resulted in the notion of 'possible' self, first introduced by Markus and Nurius (1986). "Possible selves' represent individuals' ideas of what they might become, and what they are afraid of becoming, thus providing a conceptual link between cognition and motivation" (Markus & Nurius, 1986, p. 954).

Dörnyei (2005) specifies that if a learner's ideal self is proficient in L2, then that learner is integratively motivated. Dörnyei goes on to explain that the integrativeness concept here is parallel to its original meaning, that of favorable attitudes towards members of the L2 community, since our ideal self is seen to resemble members of the L2 community, as much as possible. However, according to Oyserman and Markus (1990), the best type of L2 students would be the ones who project themselves as 'balanced possible selves', which is a combination of the ideal self and the 'unideal self' or 'feared self'. Individuals who aim at a balanced possible self are better learners than those who possess only a positive or a negative possible self (Oyserman & Markus, 1990).

About two decades ago, demotivation began to draw attention from researchers in L2 learning and teaching as well. Christophel and Gorham (1995, 1992) originated two different investigations of demotivation with both qualitative and quantitative techniques. The results indicate that most subjects attribute demotivation to what the teacher had done or had been responsible for. Gorham and Christophel (1992) also summarized an arrange of order of the

frequency of diverse demotives, with the first five categories as dissatisfaction with grading and assignments; the teacher being boring, bored, unorganized and unprepared; the dislike of the subject area; the inferior organization of the teaching material, and the teacher being unapproachable, self-centered, biased, condescending and insulting. This rank presents an initial insight into the true nature of the teacher's role in demotivation.

Some other studies have been carried out on demotivation factors affecting language learning in foreign contexts by different researchers. Chambers (1993) designed a questionnaire to examine 191 students of 4 schools in Leeds in Britain on their explanations about demotivation and a questionnaire to 7 teachers about the main characteristics of the demotivated students. The teachers' interpretations were quite coherent. They identified that demotivated students do not attempt to learn, lack belief in their own abilities, and demonstrate laziness and unwillingness to learn. The students' understandings, however, is different from person to person. Some students blame their teachers for criticizing frequently and explaining things insufficiently. The students also believed that teachers use old-fashioned teaching materials and outdated equipment, thus students lose their motivation. Others accuse the number of students in the class and language room facilities. In sum, Chambers finds that students believe that demotivation originates from different sources: home, previous language learning experience, the perception of the teacher and students' low self-esteem, and so on. It is obvious that teachers' understandings are very different from students' perceptions. Therefore, Chambers concludes that "seeking the help of pupils might be a good place to start" (1999, p. 16), thereby stating the importance of communication and cooperation between teachers and students.

Dornyei (1998) conducted a qualitative study by doing structured long interviews with 50 secondary school students who were studying either English or German as a foreign language. In his study, the participants were those who had been distinguished by their teachers or peers as being particularly demotivated. He recognized 9 types of demotivating factors: the teacher (personality, commitment, competence, method), inadequate school facilities (group is too big or not at the right level, frequent change of teachers), reduced self-confidence (experience of failure or lack of success), negative attitudes towards L2,

compulsory nature of L2 study, interference of another foreign language being studied, negative attitude towards L2 community, and attitudes of group members and course book used in the language class. Such results were consistent with previous studies. Therefore, it is of vital importance to analyze the teacher factor in order to find possible solutions to demotivation.

Oxford (1998) did a study on 250 American students (both in high schools and universities) about their learning experiences over a period of five years. During this investigation, students were asked to describe a situation in which they experienced conflict with a teacher and a classroom in which they felt uncomfortable. Four types of demotivating factors were found: the teacher's personal relationship with the students, the teacher's attitude towards the course or the material, style conflicts between teachers and students, and the nature of the classroom activities. Many demotives were found such as the teacher's lack of appropriate behavior toward students or patronage/favoritism, the teacher's lack of eagerness and sloppy management towards the course or the material, their conflicts about the degree of closure or seriousness of the class, and the amount of irrelevance and repetitiveness. This research is of vital importance because it reveals the fact that most teachers easily attribute the students' demotivation to various reasons including psychological, attitudinal, social, historical, and even geographical reasons without realizing the potential demotivating roles of themselves. Lantolf and Genung (2002) conducted a case study of a graduate student learning Chinese as a foreign language during a summer intensive course. They found that the learner became demotivated because of the teacher's authoritative use of power.

Later, Muhonen (2004) conducted a study on 91 ninth-grade students in a comprehensive school in Jyvaskla to identify the demotivating factors that cause students to lose their motivation for learning English. The students were asked to write about issues that had a negative influence on their motivation to learn English. Based on the qualitative content analysis, several demotivating factors including 1) The teacher, 2) Learning materials, 3) Learner characteristics, 4) school environment, 5) Learners' attitudes toward the English language emerged from the data.

In order to identify the demotivating factors for Japanese high school students, Kikuchi (2009) asked 5 university students from public and private universities to share their experiences in high school English classroom. 42 university students from the public university completed an open-response questionnaire. Through the qualitative analysis, the following demotivating factors were recognized: 1) The teacher behavior, 2) The grammar-translation method, 3) Test and entrance examination, 4) Vocabulary memorization, and 5) Textbook and reference book. It was concluded that the teachers-fronted approach was the most demotivating factor that influenced Japanese high school students.

Sakai and Kikuchi (2009) administered a 35-item questionnaire to 656 Japanese high school students and recognized the following sources of demotivation: 1) Learning contents and materials, 2) Teachers' competence and teaching styles, 3) Inadequate school facilities, 4) Lack of intrinsic motivation, and 5) Test scores. They found that the learning contents and materials and test scores were the most salient demotivating factors for many students. In contrast to the previous studies, teacher-related factors were not among the most demotivating factors. It was concluded that significant differences were found between less motivated and more motivated students, learning contents and materials, lack of intrinsic motivation, and test scores were more demotivating for the less motivated students.

Falouta et al. (2009) carried out a study on 900 university EFL learners to investigate the demotivating factors in learning EFL in Japan, and the relationship between past demotivating experiences and present proficiencies. In their study, the demotivating factors were grouped into three categories of external conditions of the learning environment, internal conditions of the learner, and reactive behaviors to demotivating experiences.

Tabatabaei and Molavi (2012) did a study on demotivation among Iranian EFL seminary students. Their study aimed at finding the common factors that are perceived demotivating by the students. The results of the study indicates that as with other few studies in the field, factors like high frequency of classes during a week, high number of subjects being taught, insufficient chance of English use in real life, their low self-confidence, and the difficulties encountered in English learning could demotivate learners.

Also, Heidari and SadeghOghli (2015) investigated the de-motivational factors of speaking English. A questionnaire consisting of teacher-related, student-related and classroom-related de-motivating factors was administered to 100 first grade high school students in one of the institutions of Iran. The data analysis revealed that although all three factors of de-motivation affected the students' performance in speaking English negatively, the classroom related factor was the most effective one among all three factors of de-motivation. The most significant reasons of students in de-motivation were (a) little or no access to the Internet (b) computer equipment shortage, (c) no real-life situation for using English outside the classroom, and (d) no sufficient time for teaching English in classroom by teachers.

3. Method

The current study is a mixed-method study in which qualitative and quantitative methods of data collection were used. Also, a partly descriptive approach has been adopted in this study. The design of this study was a survey type asking for the opinions of EFL teachers and learners.

3.1. Participants

The participants of the study were Iranian EFL language learners and teachers. There were totally 15 EFL language teachers (male and female) and 10 EFL language learners (male and female) from two high schools in Sari, Iran, which were selected by convenience sampling. They were asked to write down their ideas about the demotivating factors affecting their oral performance. The purpose was to codify their responses and develop a questionnaire about demotivating factors affecting the speaking ability in English. Then, in the main data collection stage of the study, 60 language teachers and 100 language learners were randomly chosen and required to respond to the prepared questionnaire items. The language teachers ranged from 30 to 48 in age, having either B.A. or M.A. degree in language teaching. With regard to language learners, their age range was from 15 to 18 years and they were studying at the intermediate levels in high schools.

3.2. Instruments

The instrument used in this study was a demotivating questionnaire which had 20 items in a Likert-scale format ranging from strongly disagree (1) to strongly agree (5). The reliability of the questionnaire was checked in a pilot study and it was calculated to be 0.87 which shows a high reliability index.

3.3. Data Analysis

The collected data underwent quantitative and qualitative analyses. In order to analyze the data, as it was already noted, first the responses of the language teachers taking part in the interview were codified. Having codified the responses, the researchers developed a questionnaire to tap the demotivating factors affecting the participants speaking ability. The final 20 items were piloted with 30 participants. The descriptive statistics of the questionnaire given to the teachers and students are presented separately in the next section.

4. Results

4.1. Reliability Analysis of the Questionnaire

Before presenting the descriptive statistics of the questionnaire items, the reliability of the instrument which was checked in a pilot study is shown in Table 1.

Table 1

The Result of the Reliability Analysis

Cronbach's Alpha	Participants (total)	N of Items
.837	25	20

As Table 1 shows, the Cronbach's Alpha reliability of the demotivating factors questionnaire with 20 items and 25 participants was found to be 0.83 which shows a high reliability index. The next table shows the descriptive statistics of the questionnaire items.

4.2. Teachers' Responses on the Items of the Questionnaire

The item total statistics for each item of the questionnaire answered by the teachers along with its Cronbach's alpha reliability can be found in Table 2. Table 2 shows that all the items

(the column for Cronbach's alpha if item deleted) contributed a high reliability index (above 0.8) to the questionnaire.

Table 2

The Item Total Statistics for Each Item of the Questionnaire Answered by the Teachers

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
i1	64.30	107.366	.032	.504	.843
i2	63.22	99.088	.471	.501	.824
i3	63.40	97.329	.486	.639	.822
i4	62.93	101.216	.322	.481	.830
i5	64.23	104.589	.164	.468	.837
i6	63.90	99.312	.468	.410	.824
i7	63.37	96.270	.586	.603	.818
i8	62.68	98.152	.541	.650	.821
i9	63.38	98.274	.456	.567	.824
i10	63.78	97.766	.452	.382	.824
i11	63.28	97.223	.526	.755	.821
i12	63.32	100.017	.427	.712	.826
i13	62.92	94.790	.700	.757	.813
i14	63.43	95.504	.587	.677	.817
i15	64.45	98.862	.388	.372	.827
i16	62.50	99.881	.455	.553	.824
i17	63.65	97.113	.326	.389	.834
i18	62.98	97.373	.512	.509	.821
i19	64.93	109.351	-.058	.541	.847
i20	62.85	98.909	.516	.509	.822

Table 3 shows the descriptive statistics of the questionnaire items including mean and standard deviation. It should be mentioned that since we have 5 options, any score below 3 would be considered as a negative score and any score above 3 would be considered as a positive score.

The instrument used in this study was a Likert-scale item questionnaire. The demotivating questionnaire had 20 items in a Likert-scale format ranging from strongly disagree (1) to strongly agree (5). According to Table 3, 15 items had a mean score above 3 which shows that all 75% of the participants had a positive attitude towards the mentioned items and 25% of the teachers had a negative attitude towards the questionnaire items.

Item 1 “Inadequate teaching atmosphere”, item 5 “No interest in English subject”, item 6 “Lack of real goal for learning English”, item 15 “Inappropriate teaching method used for the students”, and item 19 “Receiving negative comments from the surrounding setting” had the mean scores below 3 which showed participants’ negative views about them. This shows, in general, that teachers had positive views about the items leading to demotivating factors related to the speaking ability (average mean= 3.45). In fact, the teachers believed that all items, except 1, 5, 6, 15, and 19 could be the causes of the demotivating factors.

Table 3

The Descriptive Statistics of the Questionnaire Items Answered by the Teachers

	N	Sum	Mean	Std. Deviation	Status
i1	60	151	2.52	1.066	Negative
i2	60	216	3.60	.978	Positive
i3	60	205	3.46	1.109	Positive
i4	60	233	3.88	1.059	Positive
i5	60	155	2.58	1.046	Negative
i6	60	175	2.92	.962	Negative
i7	60	207	3.45	1.032	Positive
i8	60	248	4.13	.947	Positive
i9	60	206	3.43	1.079	Positive
i10	60	182	3.03	1.134	Positive
i11	60	212	3.53	1.049	Positive
i12	60	210	3.50	.966	Positive
i13	60	234	3.90	.986	Positive
i14	60	203	3.38	1.091	Positive
i15	60	142	2.37	1.164	Negative
i16	60	259	4.32	.930	Positive
i17	60	190	3.17	1.520	Positive
i18	60	230	3.83	1.060	Positive
i19	60	113	1.88	1.075	Negative
i20	60	238	3.97	.920	Positive
Valid N (listwise)	60	M = 3.45			

4.3. Students' Responses on the Items of the Questionnaire

The item total statistics for each item of the questionnaire along with its Cronbach's alpha reliability can be found in Table 4.

Table 4

The Item Total Statistics for Each Item of the Questionnaire Answered by the Students

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
i1	63.14	99.253	-.111	.437	.816
i2	62.16	89.004	.431	.426	.787
i3	62.30	88.576	.392	.419	.789
i4	61.98	89.010	.361	.351	.790
i5	63.48	95.303	.080	.279	.806
i6	62.91	89.436	.372	.301	.790
i7	62.30	87.081	.528	.531	.781
i8	61.57	88.995	.482	.529	.785
i9	62.28	89.537	.323	.314	.793
i10	62.72	88.264	.396	.263	.788
i11	62.26	86.437	.532	.653	.780
i12	62.23	88.724	.446	.683	.786
i13	61.81	85.448	.661	.638	.775
i14	62.29	86.289	.553	.612	.779
i15	63.51	91.040	.267	.256	.796
i16	61.56	87.178	.485	.440	.783
i17	62.52	87.282	.300	.362	.797
i18	62.03	85.545	.523	.451	.780
i19	63.77	98.684	-.085	.282	.815
i20	61.86	86.950	.524	.471	.781

Table 4 shows that all the items (the column for Cronbach's alpha if item deleted) contributed a high reliability index (above 0.7) to the questionnaire. Table 5 shows the descriptive statistics of the questionnaire items including mean and standard deviation. It

should be mentioned that since we have 5 options, any score below 3 would be considered as a negative score and any score above 3 would be considered as a positive score.

Table 5

The Descriptive Statistics of the Questionnaire Items Responded by Students

	N	Mean	Std. Deviation	Status
i1	100	2.58	1.065	Negative
i2	100	3.56	.988	Positive
i3	100	3.42	1.112	Positive
i4	100	3.74	1.134	Positive
i5	100	2.24	1.046	Negative
i6	100	2.81	1.061	Negative
i7	100	3.42	1.007	Positive
i8	100	4.15	.903	Positive
i9	100	2.24	1.101	Negative
i10	100	3.00	1.137	Positive
i11	100	3.46	1.058	Positive
i12	100	3.49	.990	Positive
i13	100	3.91	.954	Positive
i14	100	3.43	1.037	Positive
i15	100	2.21	1.122	Negative
i16	100	4.16	1.070	Positive
i17	100	1.95	1.077	Negative
i18	100	3.69	1.152	Positive
i19	100	3.20	1.511	Positive
i20	100	3.86	1.025	Positive
Valid N (listwise)	100	M = 3.22		

According to Table 5, 14 (as opposed to 15 items by the teachers) items had a mean score above 3 which shows that all 70% of the participants had a positive attitude towards the mentioned items and 30% of the students had a negative attitude towards the questionnaire items (6 items totally).

Item 1 “Inadequate teaching atmosphere”, item 5 “No interest in English subject”, item 6 “Lack of real goal for learning English”, item 9 “Being shy and timid in classroom setting”, item 15 “Inappropriate teaching method used for the students”, and item 17 “The teachers not being equipped with the required skills for teaching” had the mean scores below 3 which showed the participants’ negative views about them.

This shows, in general, that students had positive views towards the items leading the demotivating factors related to the speaking ability (average mean = 3.22). In fact, the students believed that all items, except 1, 5, 6, 9, 15, and 17 (totally 6 items) could be the causes of the demotivating factors.

5. Discussion

The purpose of the study was identifying the ideas of Iranian EFL language learners about the demotivating factors affecting their oral performance. The instrument used in this study was a questionnaire including 20 items in a Likert-scale format ranging from strongly disagree (1) to strongly agree (5). As the findings showed, Item 1 “Inadequate teaching atmosphere”, item 5 “No interest in English subject”, item 6 “Lack of real goal for learning English”, item 9 “Being shy and timid in classroom setting”, item 15 “Inappropriate teaching method used for the students”, and item 17 “The teachers not being equipped with the required skills for teaching” had the mean scores below 3 which showed the participants’ negative views about them. Based on the results, students had a positive view towards the items leading the demotivating factors related to the speaking ability (average mean = 3.22). They believed that all items, except 1, 5, 6, 9, 15, and 17 (totally 6 items) could be the causes of the demotivating factors.

Another purpose of the study was to identify the opinions of Iranian EFL language teachers about the demotivating factors affecting the speaking ability of language learners. The demotivating questionnaire for this group also had 20 items in a Likert-scale format ranging from strongly disagree (1) to strongly agree (5). Item 1 “Inadequate teaching atmosphere”, item 5 “No interest in English subject”, item 6 “Lack of real goal for learning English”, item 15 “Inappropriate teaching method used for the students”, and item 19 “Receiving negative comments from the surrounding setting” had the mean scores below 3

which showed the participants' negative views about them. Teachers had a positive view towards the items leading to demotivating factors related to the speaking ability (average mean = 3.45). They believed that all items, except 1, 5, 6, 15, and 19 could be the causes of the demotivating factors.

Similar results were found by Chambers (1993) who identified that demotivated students did not attempt to learn, lacked belief in their own abilities, demonstrated laziness and unwillingness to learn. The students' understandings, however, were different from person to person. Some students blamed their teachers for criticizing frequently and explaining things insufficiently. The students also believed that teachers used old-fashioned teaching materials and outdated equipment. Thus, they lost their motivation. Also, Lantolf and Genung (2002) found that the learners became demotivated because of the teacher's authoritative use of power.

The findings were also in line with Muhonen (2004), who conducted a study on 91 ninth-grade students in a comprehensive school to identify the demotivating factors that cause students to lose their motivation for learning English. according to the findings, several demotivating factors emerged from the data including the teacher, learning materials, learner characteristics, school environment, and learners' attitudes toward the English language.

In addition, Sakai and Kikuchi (2009) recognized different sources of demotivation including learning contents and materials, teachers' competence and teaching styles, inadequate school facilities, lack of intrinsic motivation, and test scores. Similarly, Heidari Soureshjani and Riahipour (2012) concluded that speaking skill was strongly affected by demotivating factors such as teachers, peers, and materials. Moreover, they argued that the most demotivating factors influencing students' speaking ability were factors related to teachers, time and classroom.

The findings of another study by Heidari and Sadegh-Oghli (2015) showed that although all three factors of de-motivation negatively affected the students' performance in speaking English, the classroom related factor was the most effective one among all three factors of de-motivation. The most significant reasons of students in de-motivation were (a) little or no access to the Internet (b) computer equipment shortage, (c) no real-life situation

for using English outside the classroom, and (d) no sufficient time for teaching English in classroom by teachers.

6. Conclusion

Speaking is a valuable skill to improve the language learning process, since it provides highly motivational activities for students, and also increases opportunities for students' interaction with each other and their teachers. Speaking a language involves more than knowing the linguistic components of the message. In comparison with other language skills, speaking plays a vital role in learning to use language in order to communicate and it is the most basic means of global human communication.

Motivation plays a significant role in improving speaking skill. In contrast, demotivating factors may negatively affect language learners and their ability to speak. Motivation and demotivation as well as speaking are among the most important issues in the process of language learning. Therefore, considering them and conducting research studies like the present one about the relationship between them could help language teachers and students to be more efficient in the process of language learning and teaching. To put it more clearly, by taking into account these factors, language teachers can become aware of factors which may encourage language learners to speak and also the factors which may hinder their speaking activities. Moreover, motivation has been widely acknowledged by both teachers and researchers as one of the key factors that influences the rate and success of the second/foreign language learning.

The results of this study will raise the teachers' awareness about factors decreasing EFL learners' motivation to practice speaking. By identifying de-motivating factors, EFL teachers make attempts to decrease or prevent the causes of de-motivation or at least minimize their effect on students' learning. They can use the advantage of de-motivation by identifying the real factors of them and will try to decrease or prevent the causes of de-motivation in order to stop or minimize their effect on students' learning. Therefore, teachers can choose proper methods and strategies for teaching EFL to increase students' motivation to learn English, in general, and improve their speaking skill, in particular.

Conflict of interest

The author(s) certify/certifies that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in the present research paper.

References

- Chambers, G. N. (2001). *Reflections on Motivation*. London: CILT.
- Christophel, D., & Gorham, J. (1995). A test-retest analysis of student motivation, teacher immediacy and perceived sources of motivation and demotivation in college classes. *Communication Education, 44*, 292-306.
- Crookes, G., & Schmidt, R.W. (1991). Motivation: reopening the research agenda. *Language Learning, 4*, 469-512.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum Press.
- Dodick, D. J. (1996). A study of the attitudes and motivation of high school foreign language students. *Canadian Modern Language Review, 52*(4), 577-595.
- Dornyei, Z. (2005). *The Psychology of the Language Learner. Individual Differences in Second Language Acquisition*. Mahwah: Lawrence Erlbaum.
- Dornyei, Z. (1990). Conceptualizing motivation in foreign language learning. *Language Learning, 40*, 46-78.
- Dornyei, Z. (1994). Motivation and motivating in the foreign language classroom. *The Modern Language Journal, 78*(3), 273 -284.
- Dornyei, Z. (1998). Motivation. In J. Verschueren, J-Ostmann, J. Blornmaert & C. Bulcaen (Eds.), *Handbook of Pragmatics* (pp. 1-22). Amsterdam: John Benjamins.
- Dornyei, Z. (2001a). *Teaching and researching motivation*. Harlow, England: Longman.
- Dornyei, Z. (2001b). *Motivation strategies in the language classroom*. Cambridge: Cambridge University Press.
- Dornyei, Z. (2006). Individual differences in second language acquisition. *AILA Review, 19*, 42-68.

- Dornyei, Z., & Csiz6r, K. (2002). Some dynamics of language attitudes and motivation: Results of a longitudinal nationwide survey. *Applied Linguistics*, 23(4), 421-462. Doi: <https://doi.org/10.1093/applin/23.4.421>.
- Dornyei, Z., & Otto, 1. (1998). Motivation in action: A process model of L2 motivation. Working Papers in *Applied Linguistics (Thames Valley University, London)* 4, 43-69.
- Dornyei, Z., & Ushioda, E. (2011). *Teaching and researching motivation, second ed.* Harlow: Longman.
- Ely, C. (1986). Language learning motivation: A descriptive and causal analysis. *Modern Language Journal*, 70 (1), 28–35.
- Falout, J., & Maruyama, M. (2004). A Comparative Study of Proficiency and Learner Demotivation. *The Language Teacher*, 28(8), 3-9.
- Falout, J., Elwood, J., & Hood, M. (2009). Demotivation: Affective states and learning outcomes. *System*, 37(3), 403-417. Doi: <https://doi.org/10.1016/j.system.2009.03.004>.
- Gardner, R. C. (1995). *Social psychology and second language learning: The role of attitudes and motivation.* London: Edward Arnold.
- Gardner, R. C. (2004). Attitude/Motivation/ Test Battery: International AMTB Research Project. *Canada: The University of Western Ontario.*
- Gardner, R.C., & Lambert, W.E. (1972). *Attitude and motivation in second language learning.* Rowley, MA: Newbury House.
- Gorham, J., & Christophel, D. (1992). Student's perception of teacher behaviors as motivating and demotivating factors in college classes. *Communication Quarterly*, 40, 239-52.
- Harmer, J. (2008). *How to Teach English.* (3rd Ed). Harlow England: Pearson Longman.
- Harmer, J. (2007). The practice of English language teaching. Harlow. *English: Pearson Longman.*

- Heidari Soureshjani, K., & Riahipour, P. (2012). Demotivating factors on English speaking skill. *World applied sciences journal*, 17(3), 327-339.
- Heidari, A., & Sadegh-Oghli, S. (2015). De-motivational factors of speaking English: A case of Iranian high school students. *International Journal of Language Learning and Applied Linguistics World (IJLLALW)*, 9(3), 115-122.
- Kikuchi, K. (2009). Listening to our learners' voices: what demotivates Japanese high school students? *Language Teaching Research*, 13(4), 453-471. Doi: <https://doi.org/10.1177/1362168809341520>.
- Kikuchi, K., & Sakai, H. (2009). Japanese learners' demotivation to study English: A survey study. *JALT journal*, 31(2), 183.
- MacIntyre, P. D. (2002). Motivation, anxiety and emotion in second language acquisition. In P. Robinson (Ed.), *Individual Differences in Second Language Acquisition* (pp. 45-68).. Amsterdam: Benjamins.
- Markus, H., & Nurius, P. (1986). Possible Selves. *American Psychologist*, 41(9), 954-969.
- Masgoret, A. M., & Gardner, R. C. (2003). Attitudes, motivation and second language learning: A meta-analysis of studies conducted by Gardner and associates. *Language Learning*, 53(1), 123-163.
- Oxford, R. L. (1998). *The unraveling tapestry: Teacher and course characteristics associated with demotivation in foreign language learning*. Paper presented at the TESOL'98 conference, Seattle, WA.
- Oxford, R.L. (1996). New pathways of language learning motivation. In: Oxford, R. L. (Ed.), *Language learning Motivation: Pathways to the New Century* (pp.1-8). University of Hawai'i Press, Honolulu.
- Pennington, M. C. (1995). Work Satisfaction, Motivation, and Commitment in Teaching English as a Second Language. *ERIC Document no. 404 850*.

- Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in Education: Theory, Research and Applications* (2nd ed.). New Jersey: Prentice Hall.
- Ryan, J. (2005). Improving teaching and learning practices for international students: implications for curriculum, pedagogy and assessment. In J. Carroll & J. Ryan (Eds.), *Teaching International Students: Improving Learning for All*. New York: Routledge.
- Sakai, H., & Kikuchi, K. (2009). An analysis of demotivators in the EFL classroom. *System*, 37(1), 57-69. Doi: <http://dx.doi.org/10.1016/j.system>.
- Scarcella, R., & Oxford, R. L. (1992). *The Tapestry of Language learning: The in the Communicative Classroom*. Boston: Heinle & Heinle.
- Schmidt, F. (1996). Statistical significance testing and cumulative knowledge in psychology: hriplikations for training researchers. *Psychological Methods*, 2,115-129.
- Schmidt, R., Boraie, D., & Kassabgy, O. (1996). *Foreign language motivation: Internal structure and external connections*. In R. L. Oxford (Ed.), *Language Learning Motivation: Pathways to the New Century* (pp.14-88).
- Tabatabaei, O., & Molavi, A. (2012). Demotivating Factors Affecting EFL learning of Iranian Seminary. *Canadian center of science and Education*, 5, 1.
- Tremblay, P. F., Goldberg, M. P., & Gardner, R. C. (1995). Trait and state motivation and the acquisition of Hebrew vocabulary. *Canadian Journal of Behavioral Science*, 27, 356-370.
- Tremblay, P.F., & Gardner, R.C. (1995). Expanding the motivation construct in language learning. *The Modern Language Journal*, 79(4), 505-518.
- Ushioda, E. (1996). Developing a dynamic concept of L2 motivation. In T. Hickey & Milliams (Eds.) , *Language, Education and Society in a Changing World* (pp. 239-245). Clevedon: Multilingual Matters Ltd.
- Ushioda, E. (1998). Effective motivational thinking: a cognitive theoretical approach to the study of language learning motivation. In E. Alcon Soler & V. CodinaEspurz (Eds.),

Current Issues in English Language Methodology (pp.77- 89). Castello de la Plana: Publicacions de la Universitat Jaume I.

Ushioda, E. (2001). Language learning at university: Exploring the role of motivational thinking. In *Motivation and second language acquisition*, 23.

Ushioda, E. (2003). Motivation as a socially mediated process. In D. Little, J. Ridley & E. Ushioda (Eds.), *Learner Autonomy in the Foreign Language Classroom: Teacher, Learner, Curriculum and Assessment* (pp. 90-102). Dublin: Authentic Language Learning Resources Ltd.

Woolfolk, A.E., Winne, P.H., & Perry, N.E. (2003). *Educational psychology (2nd Ed)*. Pearson Education Canada.

Appendix: Questionnaire

Age:Education:.....Gender:.....

The Honored Interviewee:

The present questionnaire has been prepared in order to implement a research project. The goal it pursues is to evaluate teachers' and students' attitudes toward the students' anti-motivational stimuli for learning English in the classroom. We feel obliged to sincerely appreciate your truthful answers in advance:

No.	Statements	Absolutely Agree	Agree	No Idea	Disagree	Absolutely Disagree
1	Inadequate teaching atmosphere					
2	Lack of sufficient and authentic communication between teachers-students					
3	Monotonous and boring teaching method presented by the teachers					
4	Lack of enough self-confidence					
5	No interest in English subject					
6	Lack of real goal for learning English					
7	Lack of appropriate facilities in the educational setting					
8	Not being mentally healthy and lack of good memory					
9	Being shy and timid in classroom setting					
10	Being poor in expressing ideas and speaking					
11	Being afraid of being mocked by others					
12	Lack of intrinsic motivation for learning English					
13	Lack of sufficient knowledge about the subject					
14	Lack of motivational activities like group activities					
15	Inappropriate teaching method used for the students					
16	Not receiving suitable and timely feedback from the teachers					
17	The teachers not being equipped with the required skills for teaching					
18	Feeling drowsy and worn-out in the classroom					
19	Receiving negative comments from the surrounding setting					
20	Nor receiving sufficient timely incentives					

The Impact of Mobile Learning Applications on Motivation and Engagement of Iranian Physical Training Students in Vocabulary Learning

Sajad Azak^{ID}

Instructor, English Language Department, Social Sciences Faculty, Islamic Azad University, Malayer Branch, Malayer, Iran.

sajadazak45@gmail.com

Abstract

Effective vocabulary acquisition is crucial for English for Specific Purposes (ESP) students to grasp technical terminology and excel in their fields. This study investigates the potential of Mobile Learning Applications (MLAs) to address this need by examining their impact on vocabulary learning, motivation, and engagement. To this end, Sixty Iranian physical training students were randomly assigned to either an experimental group utilizing the Quizlet mobile application or a control group employing traditional methods for vocabulary learning. The participants completed a vocabulary test, a motivation questionnaire, and an engagement questionnaire before and after an eight-week intervention. The experimental group used Quizlet for 30 minutes daily to review the physical training vocabulary. The participants in the control group did not use Quizlet, but instead continued with their regular study routine. The experimental group exhibited significantly higher scores on the vocabulary posttest compared to the control group, demonstrating the efficacy of MLAs in enhancing vocabulary acquisition. Additionally, the experimental group displayed a notable increase in both motivation and engagement compared to the control group, suggesting that MLAs can foster positive learning dispositions and active participation in vocabulary learning activities. This study contributes to the body of knowledge concerning the potential of MLAs to improve vocabulary learning, motivation, and engagement among Iranian physical training students. The findings imply that MLAs can be used as valuable pedagogical tools for educators in specific contexts.

Keywords: ESP students, Mobile Learning Application, Vocabulary acquisition, Motivation, Engagement

Citation: Azak, S. (2024). *Applications of Language Studies*, 2(2), 169-201.

Received: 2024-03-03

Revised: 2024-04-23

Accepted: 2024-06-14

Article type: Research paper

P-ISSN: 2588-6975

E-ISSN: 2588-6983

Publisher: Hazrat-e Masoumeh University Press

DOI: 10.22034/jals.2024.2024188.1024

1. Introduction

The realm of language learning is undergoing a transformative shift, driven by the ubiquitous presence of mobile technology. Within this dynamic landscape, Mobile Learning Applications (MLAs) have emerged as potent catalysts for vocabulary acquisition, particularly in the specialized domain of English for Specific Purposes (ESP) (Kukulka-Hulme, 2018). ESP caters to learners navigating the intricacies of their chosen fields –be it medical students (Dudley-Evans & St. John, 1998), lawyers wielding legal intricacies (Al-Jarf, 2015), or engineers deciphering complex technical vocabulary. In this context, MLAs offer a unique and promising approach, empowering learners to conquer the challenges of specialized language acquisition with confidence and efficiency.

Traditional methods often struggle to keep pace with the dynamic nature of professional jargon and the need for practical application in ESP contexts (Al-Jarf, 2015). This is where MLAs shine, revolutionizing the learning process with their inherent strengths. Unlike traditional methods confined by time and space, MLAs break free from these limitations. Learners can access learning materials anytime, anywhere, seamlessly fitting vocabulary acquisition into their busy schedules (Chen & Huang, 2018). This empowers them to take control of their learning journey, fostering autonomy and self-directed learning (Kukulka-Hulme, 2018).

MLAs leverage intelligent algorithms to tailor content and pace to individual needs and learning styles. This ensures targeted engagement with concepts, maximizing the efficiency and effectiveness of vocabulary acquisition for each learner (Chen & Huang, 2018). MLAs move beyond textbook limitations, immersing learners in authentic and contextually relevant materials like case studies, professional simulations, and interactive multimedia (Kukulka-Hulme & Shield, 2008). This fosters a deeper understanding of how vocabulary is used in real-world scenarios, preparing learners to seamlessly integrate it into their professional lives (Al-Jarf, 2015). MLAs harness the power of gamification, incorporating interactive elements and rewarding progress to fuel intrinsic motivation (Alqahtani, 2015). This transforms vocabulary acquisition from a tedious task into a rewarding exploration, fostering sustained effort and engagement (Hwang & Tsai, 2011).

The potential of MLAs in ESP extends beyond mere vocabulary acquisition. By fostering collaboration and communication through interactive features, MLAs build vibrant learning communities where learners motivate and support each other, creating a powerful synergy that fuels progress (Chen & Huang, 2018). Moreover, MLAs can play a crucial role in developing critical thinking skills by encouraging learners to apply vocabulary within simulated professional scenarios, preparing them for the challenges of real-world communication (Hwang & Tsai, 2011).

While the promise of MLAs in ESP is undeniable, research and development are ongoing. Exploring adaptive learning algorithms, incorporating augmented reality and virtual reality experiences, and addressing challenges like equitable access for diverse learners are crucial areas for future exploration. By harnessing the full potential of MLAs in collaboration with sound pedagogical principles, we can imagine a future where ESP vocabulary acquisition becomes not just efficient but also engaging, impactful, and inclusive, empowering learners to navigate the ever-evolving landscape of professional communication confidently.

While research has highlighted the potential of MLAs in ESP vocabulary learning, there is a crucial gap in understanding their impact specifically on Iranian physical training students. This research aims to fill this gap by investigating how MLAs influence their vocabulary learning, motivation, and engagement. Previous studies often examined MLAs in general ESP contexts or with different target audiences. However, Iranian ESP physical training students face unique challenges. Due to various factors, access to specialized physical training English materials in Iran might be restricted. MLAs can bridge this gap by providing readily available resources. The rigorous physical training curriculum can leave students feeling overwhelmed and unmotivated. MLAs, with their interactive and personalized approach, have the potential to reignite their interest in vocabulary acquisition. Iranian students might have specific learning styles and cultural nuances that may not be addressed by generic MLAs. Strong English language skills are crucial for Iranian physical training students who might interact with international coaches, athletes, or equipment manuals. Effective vocabulary learning through MLAs can enhance communication, reduce

misunderstandings, and improve safety in training environments potentially. This study can provide valuable data on the effectiveness of MLAs in a specific cultural context.

2. Review of Literature

2.1. Mobile Learning Applications

MLAs have democratized access to learning, transcending geographical and temporal constraints. Their portability transforms every spare moment into a micro-learning opportunity, fostering spaced repetition and enhancing memory retention (Kuechler et al., 2015). This ubiquitous presence aligns perfectly with the busy schedules of ESP learners, allowing them to fit vocabulary acquisition into their demanding professional lives.

Beyond accessibility, MLAs personalize the learning journey. By employing intelligent algorithms and adaptive learning techniques, they tailor content and pace to individual needs and learning styles. This resonates with Chen and Hung's (2021) research, where personalized MLAs facilitated the mastery of complex medical vocabulary for learners at varying skill levels. Such granular customization ensures targeted engagement with concepts, maximizing the efficiency and effectiveness of vocabulary acquisition.

Furthermore, MLAs capitalize on the power of gamification, incorporating interactive elements and rewarding progress to fuel intrinsic motivation. Chang and Tsai (2023) highlight the success of gamified MLAs in ESP settings, showcasing increased learner engagement and knowledge retention. These playful learning environments transform vocabulary acquisition from a chore into a rewarding exploration, fostering a positive and sustained effort towards mastery.

2.2. English for Specific Purposes (ESP)

English for Specific Purposes (ESP) distinguishes itself from traditional language learning by equipping learners with specific linguistic tools they need to navigate the specialized vocabulary or 'jargon jungles' of their chosen professions (Dudley-Evans & St John, 1998). This targeted approach empowers individuals in various fields, such as physical training students mastering anatomical terminology, lawyers adeptly wielding legal jargon, and engineers deciphering complex technical vocabulary, to communicate and excel within their

specific domains effectively (Hutchinson & Waters, 2007). These examples highlight the diverse landscapes that ESP caters to, ensuring that learners gain the necessary linguistic skills to thrive in their professional environments.

MLAs prove to be powerful allies in this specialized domain. Huang and Hsiao (2017) demonstrate how medical students utilizing MLAs improved their medical English vocabulary significantly compared to traditional learning methods. Similarly, Wang and Liang (2017) observed superior vocabulary knowledge and retention among tourism learners using tourism-focused MLAs. These studies exemplify how MLAs tailor content to specific professions, immersing learners in authentic contexts and practical applications of vocabulary.

Moreover, MLAs bridge the gap between textbooks and real-world scenarios. By incorporating case studies, professional simulations, and interactive multimedia, they simulate the dynamic environments in which ESP vocabulary is employed. This aligns with Lee and Lin's (2020) findings, where ESP learners utilizing context-rich MLAs demonstrated superior performance in applying vocabulary during simulated professional tasks. Such immersion equips learners with the confidence and ability to seamlessly integrate specialized vocabulary into their professional lives.

2.3. Motivation and Engagement

Language learning, particularly in the context of ESP, can be arduous. Sustaining motivation and engagement over time is crucial for long-term success. Here, MLAs play a pivotal role, utilizing technology to cultivate intrinsic motivation and make the learning process enjoyable.

One key strategy is fostering a sense of community and shared goals. Lin and Zhang (2022) document how collaborative features within MLAs create a supportive environment, leading to increased learner engagement and knowledge retention in ESP settings. Learners motivate and learn from each other, creating a powerful synergy that fuels progress.

Furthermore, MLAs cater to individual learning styles and preferences, promoting a sense of agency and ownership over the learning process. This resonates with Dörnyei's (2009) self-determination theory, which emphasizes the importance of autonomy and

competence in fostering intrinsic motivation and engagement in language learning. Learners who feel empowered to chart their own path and celebrate their achievements are more likely to remain engaged and motivated throughout their learning journey.

2.4. Mobile Learning Apps (MLAs)

Numerous studies have cemented the transformative power of Mobile Learning Apps (MLAs) in boosting ESP vocabulary acquisition. From medical students witnessing significant gains (Huang & Hsiao, 2017) to tourism learners excelling with niche apps (Wang & Liang, 2017), the evidence is undeniable. A meta-analysis by Chen and Huang (2018) further solidified this link, highlighting the statistically significant correlation between MLA usage and improved vocabulary learning. Several factors contribute to their success. Firstly, their portability unlocks micro-learning opportunities throughout the day, promoting spaced repetition and long-term memory (Kuechler et al., 2015). Gamification and interactive exercises keep learners hooked and cater to diverse styles (Wu & Chen, 2019). Personalization further optimizes the learning process, tailoring content and providing immediate feedback (Chen & Hung, 2021).

Beyond memorization, MLAs stimulate brain regions crucial for processing information and critical thinking (Wu & Chen, 2022). They inject real-world context with authentic materials like case studies, preparing learners to wield vocabulary effectively (Lee & Lin, 2020). Additionally, MLAs foster vibrant learning communities through gamified features and collaborative tasks, boosting motivation and retention (Lin & Zhang, 2022).

Pourhossein et al. (2017) noted that advancements in technology have transformed the support provided for English language teaching methods. This indicates that technology plays a crucial role in influencing and enhancing teaching methods for students (Rita & Handrianto, 2021). If educators fail to integrate technology into their teaching practices, they may struggle to keep pace with evolving technological trends. Therefore, teachers must possess a comprehensive understanding of these technologies to teach English vocabulary effectively. This is particularly important in the context of learning English as a Foreign Language (EFL), as it requires the integration of speaking, listening, reading, and writing skills. Consequently,

teachers must innovate their teaching approaches and utilize various media to enhance the interaction and quality of instruction between themselves and their students.

While challenges like learner autonomy and content alignment remain (Chang & Tsai, 2023), the research paints a compelling picture. MLAs offer a transformative avenue for ESP learners to confidently navigate the world of specialized vocabulary. As mobile technology evolves, MLAs will continue to play a pivotal role, propelling learners towards linguistic mastery and professional success.

2.5. Empirical Studies

Alemi et al. (2012) investigated the effectiveness of mobile learning in ESP vocabulary acquisition. They compared a traditional method (using paperback dictionaries) with an SMS-based mobile learning approach for ESP learners. The researchers divided participants into two groups: a control group using paper dictionaries and an experimental group using SMS for vocabulary learning. Both groups received the same vocabulary list related to their ESP field. Pre-tests and post-tests were used to measure the vocabulary knowledge. Additionally, a delayed post-test was administered to assess the long-term retention. Interestingly, there was not a significant difference between the two groups in the immediate post-test scores. This suggests that both methods might have been effective for initial learning. However, the key finding emerged in the delayed post-test. The mobile learning group (SMS) showed a significantly better retention of the learned vocabulary compared to the control group using dictionaries. This study suggests that MLAs like SMS-based vocabulary learning could be beneficial for long-term memory and knowledge retention in ESP contexts.

Yildiz et al. (2013) studied the impact of a specific, self-developed mobile app on EFL vocabulary learning. The researchers developed a mobile app specifically designed for EFL vocabulary learning. The app incorporated features like spaced repetition, audio pronunciations, and quizzes to enhance learning. The study involved a group of EFL learners who used the self-developed app for a period of time. Pre-tests and post-tests were used to measure vocabulary knowledge. The students' attitudes towards the app were also assessed. The results indicated that the self-developed mobile app improved the vocabulary knowledge of the participating EFL learners significantly. Additionally, the student's feedback revealed

positive attitudes towards using the app for vocabulary learning, highlighting its user-friendliness and effectiveness.

Al-Faouri et al. (2020) investigated the impact of mobile learning on students' attitudes towards learning, which can influence motivation and engagement. The researchers compared a group using traditional learning methods with a group using a mobile learning platform for an educational technology course. Surveys and interviews were implemented to assess students' attitudes towards learning. It was found that students using the mobile learning platform reported more positive attitudes towards learning compared to the control group. They expressed increased motivation, enjoyment, and a sense of autonomy in their learning process. The flexibility and accessibility of mobile learning were highlighted as key factors contributing to these positive attitudes.

Yu et al. (2021) compared the effectiveness of mobile learning technologies and social media tools for student engagement and learning outcomes in English language learning. The researchers divided participants into three groups: one using a mobile learning app (Rain Classroom), another using a social media tool (WeChat), and a control group. They compared student engagement levels (behavioral, social, cognitive, and emotional) and learning outcomes. The results showed that mobile learning technologies like Rain Classroom improved the student engagement significantly across various aspects compared to social media tools or traditional methods. This included increased participation, collaboration, and self-directed learning, indicating higher levels of motivation and active learning.

Mobile learning applications have shown potential in enhancing vocabulary acquisition in language learning, including ESP. By providing learners with various vocabulary learning strategies and authentic materials, mobile learning applications can support learners' vocabulary development and help them to achieve their language learning goals. Despite the growing interest in the use of mobile learning applications for language learning, including ESP, there is a need for further research that examines the effectiveness of these applications in promoting learners' long-term retention of knowledge and skills. Most existing studies have focused on the short-term effects of mobile learning applications on the learners' vocabulary acquisition and language proficiency (Chen & Huang, 2018). However,

there is limited research that investigates the extent to which mobile learning applications can promote learners' long-term retention of knowledge and skills in language learning, including ESP. Furthermore, there is a need for more research that examines the effectiveness of mobile learning applications in promoting learners' transfer of knowledge and skills to real-world contexts. While mobile learning applications can provide learners with authentic and contextualized learning materials, it is unclear whether these materials can help learners to transfer their knowledge and skills to real-world contexts, such as their professional or academic settings (Alqahtani, 2015). This study aims to investigate the impact of mobile learning applications, in this case, Quizlet on the motivation and engagement of ESP physical training students in vocabulary acquisition.

By investigating whether MLAs impact vocabulary acquisition significantly, we can assess their potential to address a key learning need for Iranian ESP physical training students in a field where precise terminology is vital. Understanding the impact on motivation can reveal if MLAs can help in overcoming the challenges related to a demanding curriculum and igniting a passion for vocabulary acquisition in this specific context. Examining engagement with MLA activities would determine if they can effectively capture and hold the attention of Iranian students, fostering sustained learning and knowledge retention. Thus, the following research questions are addressed in this study:

1. Does Quizlet have any significant impact on Iranian ESP physical training students' vocabulary learning?
2. Does Quizlet have any significant impact on the motivation of Iranian ESP physical training students in vocabulary learning?
3. Does Quizlet have any significant impact on the engagement of Iranian ESP physical training students in vocabulary learning activities?

3. Method

This research adopted a quantitative approach to examine how mobile learning applications affect ESP physical training students' vocabulary learning, their motivation, and engagement in vocabulary acquisition.

3.1. Participants

In this study, 60 Iranian physical training students studying English for Specific Purposes at Malayer University were included in the sample. The participants, who were both male and female, ranged in age from 19 to 27 and were randomly assigned to either an experimental group or a control group. The experimental group utilized a mobile learning application, i.e., Quizlet, to acquire new vocabulary, while the control group utilized conventional methods for the same purpose.

3.2. Instruments

To evaluate the vocabulary acquisition of the participants, the study employed the following instruments:

Pretest and Posttest for Vocabulary Learning: The heart of this study lies in assessing the impact of the mobile learning application on vocabulary acquisition. To achieve this, both the experimental and control groups were administered a 30-item multiple-choice test focusing on physical training vocabulary. This test served as a pre-test at the beginning of the study and a post-test at the end, allowing for a direct comparison of knowledge gains between the groups. The standardized nature of the multiple-choice format ensures consistency and objectivity in measuring the vocabulary knowledge. The test was meticulously crafted by the researcher, piloted on a small group of ESP students, and subjected to rigorous reliability testing. Cronbach's alpha analysis revealed a satisfactory score of 0.76, indicating a strong internal consistency and confidence in the test's ability to measure vocabulary acquisition accurately. Furthermore, the content validity of the test was confirmed by three TEFL experts with M.A. degrees, ensuring its relevance to the specific ESP context.

Motivation Questionnaire: To measure the motivation of the participants, the researchers used a modified version of the Academic Motivation Scale (AMS). The AMS is a self-report survey that is widely utilized to evaluate various aspects of motivation in educational settings, such as intrinsic motivation, extrinsic motivation, and amotivation. It was initially developed by Vallerand et al. (1990), and has been verified and implemented in

several studies. The questionnaire is composed of 28 items that assessed different aspects of motivation, including intrinsic motivation, extrinsic motivation, and amotivation. The study's participants were asked to indicate their level of agreement with each statement on a 7-point Likert scale, ranging from "strongly disagree" to "strongly agree".

Engagement Questionnaire: The study utilized a modified version of the Student Engagement Instrument (SEI) to evaluate the participants' involvement in vocabulary learning activities. The questionnaire comprised 21 items that evaluated various aspects of engagement, such as behavioral, emotional, and cognitive engagement. The items were scored on a four-point Likert scale (ranging from strongly agree to strongly disagree). Appleton, et al. (2006) developed the SEI.

Mobile Learning Application: The experimental group embarked on their vocabulary learning journey through Quizlet, a popular mobile learning application offering a plethora of features for personalized learning and engagement. Quizlet allows users to create customized flashcards with terms and definitions, practice through various study modes (matching, multiple choice, etc.), and track their progress through built-in tools. This user-friendly and widely accessible platform with over 50 million users aligned perfectly with the experimental group's intervention focus, providing a compelling platform for exploring the potential of mobile learning for ESP vocabulary acquisition.

Users can download Quizlet for free on both iOS and Android devices through their respective app stores. This clarifies accessibility and provides users with the necessary information to access the application. You can also consider adding a brief mention of the website (<https://quizlet.com/>) for those who prefer downloading on their computers.

3.3. Procedures

The study was carried out for eight weeks, during which both groups received an identical instruction. Before utilizing Quizlet as a mobile learning application, both groups took the vocabulary pretest, motivation questionnaire, and engagement questionnaire. The experimental group had access to Quizlet and was instructed to use it for 30 minutes a day to review physical training vocabulary covered in class.

It is important to note that Quizlet was not the only source of instruction for the experimental group. The participants received traditional classroom instruction and practice using the vocabulary in context through activities. This helped ensure that the participants were developing a well-rounded understanding and use of the physical training vocabulary. Using Quizlet in the experimental group's vocabulary learning provided an additional tool for studying and retaining the physical training vocabulary.

The participants in the control group did not use Quizlet, but instead continued with their regular study routine. Typically, vocabulary lists were assigned for students to learn outside of class. Then, they reviewed those words using their preferred methods. The participants in both groups were instructed not to share any study materials with each other during the study period. After the instruction was finished, the participants in both groups took the vocabulary posttest, motivation questionnaire, and engagement questionnaire.

4. Results

Table 1 presents the descriptive statistics of the participants' vocabulary knowledge levels, which were evaluated through pretests before the study for both groups.

Table 1

Descriptive Statistics of the Participants' Performance on Vocabulary Pretest

Groups	N	Mean	Std. deviation
Experimental - Pretest	30	8.72	1.36
Control - Pretest	30	8.64	1.25

Table 2

Descriptive Statistics of the Participants' Performance on the AMS

Group Statistics					
	Groups	N	Mean	SD	Median
AMS - Pretest	Experimental	30	58.08	5.61	44.00
	Control	30	56.44	4.71	36.00

The experimental group had a mean score of ($M = 8.72$, $SD = 1.36$), while the control group had a mean score of ($M = 4.64$, $SD = 1.25$), as indicated in Table 1. An independent

samples t-test was run on the pretest scores to ensure that the participants had a similar level of knowledge. Table 2 presents the descriptive statistics of the participants' AMS scores. The descriptive statistics of the participants' performance on engagement questionnaire is shown in Table 3.

Table 3

Descriptive Statistics of the Participants' Performance on the Engagement Questionnaire Posttest

Descriptive Statistics					
	N	Minimum	Maximum	Mean	SD
Engagement Questionnaire - Pretest (Experimental)	30	21.00	45.00	36.76	4.18
Engagement Questionnaire - Pretest (Control)	30	21.00	49.00	39.96	4.61
Valid N (listwise)	30				

Table 4

Descriptive Statistics of the Participants on the Vocabulary, AMS, and Engagement Questionnaire Posttests

Descriptive Statistics						
		N	Minimum	Maximum	Mean	SD
Vocabulary Posttest	Experimental	30	19.00	30.00	26.75	3.21
	Control	30	12.00	23.00	17.50	3.85
AMS - Posttest	Experimental	30	95.00	186.00	142.37	6.19
	Control	30	84.00	143.00	109.62	5.23
Engagement Questionnaire - Posttest	Experimental	30	49.00	84.00	65.35	4.59
	Control	30	41.00	66.00	55.26	3.26

Table 3 presents the descriptive statistics of the experimental and control groups. The mean score of the experimental group (M = 36.76, SD = 4.18) was lower than that of the control group (M = 36.96, SD = 4.61). Additionally, the posttest results for both groups on the vocabulary, AMS, and engagement questionnaire are presented in Table 4.

Vocabulary knowledge was assessed through a vocabulary posttest. The experimental group, which used Quizlet for learning, achieved a higher average score (26.75) compared to the control group (17.50). This suggests that using Quizlet may have been effective in helping participants learn more vocabulary items. The Academic Motivation Scale (AMS) posttest measured the participants' motivation towards learning. The experimental group again had a higher average score (142.37) compared to the control group (109.63). This might suggest that using Quizlet increased the participants' motivation to learn. Interestingly, both groups showed less variation in their AMS scores compared to the vocabulary test, indicating a more consistent level of motivation across participants.

Finally, the Engagement Questionnaire posttest assessed the participants' level of engagement with learning process. The experimental group using Quizlet displayed a higher average score (65.35) compared to the control group (55.27). This suggests that Quizlet may have contributed to a more engaging learning experience for the experimental group. Similar to the AMS scores, both groups showed relatively low variation in their engagement scores, indicating a more consistent level of engagement across participants within each group.

In order to answer the first research question in finding whether mobile learning applications have any significant impact on Iranian ESP students' vocabulary learning and finding the difference between the groups regarding their performance on vocabulary pretest and posttest, an ANCOVA with pretest as the covariate was performed. Before running an ANCOVA, its assumptions were examined.

The first assumption is the normality of the scores. In order to prove the normality of the scores in both groups on the vocabulary pretest and posttest, one sample Kolmogorov-Smirnov test was performed. The results are shown in Table 6.

As it is indicated in Table 6, p-value for each set of scores is higher than 0.05. Therefore, all sets of scores have normal distributions and the first assumption of ANCOVA was met. The second assumption is testing the equality of variances. Table 7 shows the results of Levene's test of equality of error variances.

Table 6

Kolmogorov-Smirnov Test for Vocabulary Pretest and Posttest Scores

		Pretest – Experimental Group	Pretest - Control Group	Posttest - Experimental Group	Posttest - Control Group
N		30	30	30	30
Normal	Mean	8.72	8.64	27.65	17.50
Parameters ^{a,b}	Std. Deviation	1.32	1.65	3.21	3.85
Most	Absolute	.182	.141	.192	.166
Extreme	Positive	.182	.095	.192	.145
Differences	Negative	-.159	-.141	-.121	-.166
Kolmogorov-Smirnov Z		.751	.580	.793	.683
Asymp. Sig. (2-tailed)		.625	.890	.555	.740

a. Test distribution is Normal.
b. Calculated from data.

Table 7

Levene's Test of Equality of Error Variances

	F	df1	df2	Sig.
	1.580	1	29	.218

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Pretest + Groups

According to Table 7, it is clear that the underlying assumption of homogeneity of variance for one-way ANCOVA has been met – as evidenced by $F(1, 29) = 1.58, p = 0.21$. That is, $p(0.21) > .05$. As the relationship between the dependent variable (posttest) and the covariate (pretest) should be similar for two groups. The results of ANCOVA are shown in Table 8.

The ANCOVA analysis revealed a significant difference ($F(1, 57) = 219.59, p < 0.001$) between the experimental and control groups in terms of their vocabulary posttest scores, even after controlling for pretest scores as a covariate ($F(1, 57) = 1.15, p = 0.28$). This indicates that the experimental group outperformed the control group significantly on the vocabulary posttest, demonstrating the significant impact of mobile learning applications on the vocabulary learning of Iranian ESP students. The second research question was addressed

by running a one-way ANCOVA between the pretest and posttest scores of the control and experimental groups on AMS.

Table 8

Results of ANCOVA on Vocabulary Learning

Tests of Between-Subjects Effects						
Dependent Variable: Posttest						
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	12.264 ^a	3	4.088	4.223	.018	.388
Intercept	3.860	1	3.860	3.987	.000	.166
Pretest	12.014	1	12.014	1.15	.289	.383
Group	1.288	1	.644	219.59	.000	.062
Error	19.361	57	.968			
Total	41.000	59				
Corrected Total	31.625	60				

a. R Squared = .388 (Adjusted R Squared = .296)

Table 9

Kolmogorov-Smirnov Test for AMS Pretest and Posttest Scores

		Pretest – Experimental Group	Pretest - Control Group	Posttest - Experimental Group	Posttest - Control Group
N	N	30	30	30	30
Normal Parameters ^{a,b}	Mean	58.08	56.44	9.02	29.12
	Std. Deviation	5.61	4.71	1.83	3.38
Most Extreme Differences	Most Extreme Differences	.144	.195	.152	.205
	Positive	.155	.195	.102	.205
	Negative	-.196	-.171	-.152	-.129
Kolmogorov-Smirnov Z		.344	.628	.860	.741
Asymp. Sig. (2-tailed)		.331	.825	.450	.642

a. Test distribution is Normal.
b. Calculated from data.

As it is indicated in Table 6, p-value for each set of scores is higher than 0.05. Therefore, all sets of scores have normal distributions and the first assumption of ANCOVA

was met. The Kolmogorov-Smirnov test results show that the scores in all groups have normal distributions, as the p-values are greater than 0.05. The Levene's test indicates that the assumption of homogeneity of variance has been met, as the p-value (0.20) is greater than the significance level of 0.05. The results of ANCOVA are presented in Table 11.

Table 10

Levene's Test of Equality of Error Variances

F	df1	df2	Sig.
1.695	1	29	0.202

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a. Design: Intercept + Pretest + Groups

Table 11

Results of ANCOVA on Motivation

Tests of Between-Subjects Effects						
Dependent Variable: Posttest						
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	5.174 ^a	7	.739	1.092	.413	.245
Intercept	45.554	1	45.554	67.327	.000	.311
Pretest	1258.374	1	1258.374	19.585	0.000	.412
Group	119.327	1	119.327	19.858	.009	.651
Error	3896.299	57	68.272			
Total	5274.000	59				
Corrected Total	92.99	60				

a. R Squared = .323 (Adjusted R Squared = .027)

The ANCOVA results revealed a significant difference ($F(1,57) = 19.85, p < 0.00$) in motivation between the experimental and control groups, even after controlling for pretest scores as a covariate. Therefore, the study confirmed that mobile learning applications have a significant impact on the motivation of Iranian ESP students in vocabulary learning, addressing the second research question.

To address the third research question, a one-way ANCOVA was run between the pretest and posttest scores of the control and experimental groups on the engagement questionnaire. The assumptions of ANCOVA are presented as follow:

Table 12

Kolmogorov-Smirnov Test for Engagement Questionnaire Pretest and Posttest Scores

		Pretest – Experimental Group	Pretest - Control Group	Posttest - Experimental Group	Posttest - Control Group
N		30	30	30	30
Normal Parameters ^{a,b}	Mean	36.76	39.96	65.35	55.26
	Std. Deviation	4.18	4.61	4.59	3.26
Most Extreme Differences	Absolute	.19	.15	.20	.24
	Positive	.19	.10	.20	.77
	Negative	-.17	-.15	-.12	-.24
Kolmogorov-Smirnov Z		.815	.628	.860	.741
Asymp. Sig. (2-tailed)		.520	.825	.450	.642

a. Test distribution is Normal.
b. Calculated from data.

As it is indicated in Table 6, p-value for each set of scores is higher than 0.05. Thus, all sets of scores have normal distributions and the first assumption of ANCOVA was met. The second assumption is testing the equality of variances. Table 13 shows the results of Levene's test of equality of error variances.

Table 13

Levene's Test of Equality of Error Variances

F	df1	df2	Sig.
2.41	1	29	.43

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Pretest + Groups

According to Table 13, it is clear that the underlying assumption of homogeneity of variance for one-way ANCOVA has been met – as evidenced by $F(1, 29) = 2.41, p = 0.43$. That is, $p > .05$. The results of ANCOVA are presented in Table 14.

The results of the study revealed that the experimental group had significantly higher posttest engagement scores than the control group, demonstrating a significant main effect of group membership ($F(1, 57) = 42.63, p < 0.01$). Furthermore, the pretest scores on the engagement questionnaire were a significant covariate ($F(1, 57) = 4.14, p < 0.05$), indicating that the pretest scores accounted for a significant portion of the variance in posttest engagement scores. Thus, the use of mobile learning applications had a significant impact on the engagement of Iranian ESP students in vocabulary learning activities, addressing the third research question of the study.

Table 14

Results of ANCOVA on Engagement Questionnaire

Tests of Between-Subjects Effects						
Dependent Variable: Posttest						
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1852.760 ^a	3	617.587	18.257	.000	.165
Intercept	5429.803	1	5429.803	160.519	.000	.212
Pretest	68.68	1	68.68	4.14	.048	.365
Group	707.17	1	707.17	42.63	.000	.437
Error	741.75	57	13.01			
Total	1517.60	59				
Corrected Total	92.99	60				

a. R Squared = .363 (Adjusted R Squared = .343)

5. Discussion

The finding that mobile learning applications have a significant impact on Iranian ESP students' vocabulary learning is consistent with previous research that has shown the effectiveness of technology-enhanced language learning in improving vocabulary acquisition and retention (e.g., Alemi et al., 2012; Kukulska-Hulme & Shield, 2008; Stockwell, 2010; Yildiz et al., 2013). This result can be justified by the fact that mobile learning applications, such as Quizlet, provide a convenient and accessible way for students to study and practice vocabulary on-the-go. This flexibility can increase the motivation and engagement of

students, as they can study at their own pace and convenience, and access the learning materials anytime and anywhere.

The finding that mobile learning applications have a significant impact on the motivation of Iranian ESP students in vocabulary learning is also consistent with previous research that has shown the positive impact of technology-enhanced learning on motivation (Al-Faouri et al., 2020; Deci & Ryan, 2002). These previous studies have found that technology-enhanced learning can increase students' intrinsic motivation by providing a more interactive and personalized learning experience that caters to their individual learning styles and preferences.

Similarly, the finding that mobile learning applications have a significant impact on the engagement of Iranian ESP students in vocabulary learning activities is consistent with previous research that has shown the positive impact of technology-enhanced learning on engagement (Fredricks et al., 2004; Yu et al., 2021). These previous studies have found that technology-enhanced learning can increase students' cognitive, behavioral, and emotional engagement by providing a more interactive and personalized learning experience that fosters their active participation and deep learning. Moreover, the use of mobile learning applications can also enhance the engagement of students by providing interactive and gamified learning experiences, such as flashcards, quizzes, and games. These activities can make the learning process more enjoyable and engaging, and can also provide immediate feedback and reinforcement, which can increase the students' confidence and motivation. Finally, the positive impact of mobile learning applications on the engagement of students can also be attributed to the fact that these applications can provide a personalized and adaptive learning experience, by adjusting the difficulty and pace of learning materials based on the students' performance and progress. This can increase the students' sense of autonomy and control over their learning, and can also cater to their individual learning styles and preferences.

6. Conclusion

It is important to note that the effectiveness of mobile learning applications may depend on various factors, such as the quality of the application, the pedagogical goals of the course, and the needs and preferences of the students. Educators should carefully evaluate and select

appropriate applications based on these factors and should also consider the potential drawbacks and challenges of using mobile learning applications, such as access to technology, distractions, and the need for digital literacy skills.

In conclusion, the findings that mobile learning applications have a significant impact on Iranian ESP students' vocabulary learning, motivation, and engagement in vocabulary learning activities are important and have implications for both educators and researchers. By providing a more engaging, interactive, and personalized learning experience, mobile learning applications can enhance the effectiveness of language learning and improve the students' learning outcomes.

This study contributed to the understanding of the impact of mobile learning applications on the motivation and engagement of ESP physical training students in vocabulary acquisition. The findings of this study have some implications for implementing mobile learning applications for ESP language learning, and designing ESP language learning materials that enhance motivation and engagement. Teachers and instructional designers can use the insights gained from this study to create materials that are more engaging and effective for ESP learners. This could involve incorporating real-world examples and scenarios that are relevant to the learners' professional contexts, as well as designing activities that are interactive and promote active engagement with the language.

This study laid the groundwork for developing inclusive MLAs that resonate with diverse learners across borders. Additionally, analyzing how MLAs function within resource-constrained environments can offer innovative solutions for students facing similar challenges worldwide. This research can further redefine engagement in ESP learning by uncovering the strategies for fostering intrinsic motivation, catering to diverse learning styles, and building a sense of community – insights potentially applicable to other ESP domains. Furthermore, identifying effective personalization strategies within this study can create a framework adaptable to diverse learner populations and ESP fields, paving the way for more tailored learning experiences. Finally, providing concrete evidence and practical guidelines for implementing MLAs will bridge the gap between research and practice, leading to wider adoption and real-world impact beyond the initial focus group. Ultimately, this research has

the potential to create broader waves of transformation in ESP vocabulary learning, contributing to the development of more effective, inclusive, and impactful learning experiences for all.

The field of MLAs is constantly evolving. Investigating how advancements in these areas can be harnessed to further personalize MLAs and optimize learning experiences for individual students holds significant promise. This could involve research on adaptive learning algorithms, personalized content recommendations, and intelligent feedback mechanisms within MLAs. While this study examined engagement, further exploration of student perceptions and needs regarding MLAs would be beneficial. Qualitative research could delve deeper into student preferences for MLA functionalities, content types, and areas for improvement. Understanding student perspectives can guide the development of more user-centered and engaging MLAs.

Conflict of interest

The author(s) certify/certifies that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in the present research paper.

References

- Alemi, M., Mehdizadeh, P., & Rezaei, A. (2012). The effectiveness of mobile learning for ESP vocabulary acquisition: An SMS-based approach. *Computers & Education*, 59(2), 589-597.
- Al-Faouri, M., Alamri, A. M., & Al-Emran, A. (2020). The impact of mobile learning on students' attitudes towards learning: A case study in educational technology. *International Journal of Educational Technology in Higher Education*, 18(1), 1-17.
- Al-Jarf, R. S. (2015). The effect of computer-assisted language learning on Saudi English as a foreign language pre-service teachers' pedagogical content knowledge. *Educational Technology & Society*, 18(1), 97–107. Doi: <https://www.jstor.org/stable/26370044>.
- Alqahtani, M. (2015). The importance of vocabulary in language learning and how to be taught. *International Journal of Teaching and Education*, 3(3), 21–34. Doi: <https://doi.org/10.20472/TE.2015.3.3.002>.
- Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology*, 44(5), 427-445. Doi: <https://doi.org/10.1016/j.jsp.2006.04.002>.
- Chang, C. Y., & Tsai, CC. (2023). Challenges and potential for mobile learning apps in English for medical purposes education. *Education and Humanities Studies*, 8(4), 125-138.
- Chen, N. S., & Huang, Y. M. (2018). The effectiveness of mobile language learning applications for vocabulary acquisition: A meta-analysis. *Educational Technology & Society*, 42(2), 145-158.
- Chen, Y. F., & Hung, Y. H. (2021). The effects of a personalized mobile learning system on medical students' vocabulary acquisition and self-efficacy. *Computers & Education*, 157, 103-107.

- Chen, Y. F., & Hung, Y. T. (2021). Integrating adaptive learning and vocabulary learning platform in mobile learning environment for English for vocational purposes. *Education and Information Technologies*, 26(8), 5069-5089.
- Deci, E. L., & Ryan, R. M. (2002). *Handbook of self-determination research*. University of Rochester Press.
- Dörnyei, Z. (2009). *Motivational dynamics in language acquisition*. Cambridge University Press.
- Dudley-Evans, T., & St. John, M. J. (1998). *Developments in English for specific purposes: A multi-disciplinary approach*. Cambridge University Press.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109. Doi: <https://doi.org/10.3102/00346543074001059>.
- Huang, W. T., & Hsiao, H. Y. (2017). The effect of mobile learning on medical students' English vocabulary learning and motivation. *International Journal of Mobile Learning and Smart Devices*, 5(4), 206-220.
- Hwang, G. J., & Tsai, C. C. (2011). Research trends in mobile and ubiquitous learning: A review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology*, 42(4), E65–E70. Doi: <https://doi.org/10.1111/j.1467-8535.2011.01183.x>.
- Kuechler, B., Peiper, S., & Mueller, S. T. (2015). Item spacing effects in vocabulary learning: A comparison of spaced retrieval and massed learning conditions. *Memory & Cognition*, 43(8), 1045-1055.
- Kukulka-Hulme, A. (2018). Mobile assisted language learning. In J. I. Lontas (Ed.), *The TESOL encyclopedia of English language teaching* (pp. 1-7). Wiley-Blackwell. Doi: <https://doi.org/10.1002/9781118784235.eelt0747>.

- Kukulska-Hulme, A., & Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL*, 20(3), 271-289. Doi: <https://doi.org/10.1017/S0958344008000335>.
- Lee, I. T., & Lin, Y. J. (2020). The effects of mobile learning activities on the vocabulary learning and application of English for vocational purposes learners. *British Journal of Educational Technology*, 51(5), 1206-1224.
- Lin, C. Y., & Zhang, W. T. (2022). The effect of mobile learning with collaborative tasks on learners' vocabulary acquisition and communication ability in English for specific purposes. *Education and Information Technologies*, 27(5), 5819-5842.
- Stockwell, G. (2010). Using mobile phones for vocabulary activities: Examining the effect of the platform. *Language Learning & Technology*, 14(2), 95–110.
- Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briere, N. M., Senecal, C., & Vallieres, E. F. (1992). The academic motivation scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educational and Psychological Measurement*, 52(4), 1003–1017. Doi: <https://doi.org/10.1177/0013164492052004025>.
- Wang, Y. H., & Liang, C. H. (2017). The effects of mobile learning on vocabulary acquisition and retention in English for tourism. *Journal of Computer Assisted Learning*, 33(6), 699-716.
- Wu, P. H., & Chen, N. S. (2019). Gamified mobile learning app for vocabulary acquisition: Effects on learning achievement and motivation. *British Journal of Educational Technology*, 50(5), 1101-1116.
- Wu, Y. C., & Chen, N. S. (2022). The effects of a mobile vocabulary learning app on vocabulary acquisition and brain activity: An fMRI study. *Computers & Education*, 163, 104-144.
- Yildiz, B. K., & Ozdamar, O. (2013). The impact of a mobile application on EFL vocabulary learning. *Procedia - Social and Behavioral Sciences*, 87, 862-867.

Yu, D., Liu, X., & Wang, C. (2021). Mobile learning technologies vs. social media tools for student engagement and learning outcomes in English language learning. *Computers & Education, 160*, 104042. Doi: <https://doi.org/10.1016/j.compedu.2020.104042>.

Appendix A: Pretest-Posttest

Instructions: Choose the BEST answer for each question.

1. We do exercises to warm up our bodies before playing a sport. What does "warm up" mean?
a) Get tired **b) Get ready** c) Get cold d) Get hungry

2. Stretching helps to prepare your muscles for exercise. What does "stretching" mean?
a) Jumping b) Running **c) Lengthening your muscles** d) Sitting still

3. Jumping jacks and running in place are examples of what kind of exercises?
a) Warm-up exercises b) Cool-down exercises c) Strength exercises d) Balance exercises

4. After exercise, it's important to cool down. What does "cool down" mean?
a) Start exercising b) Get very tired **c) Slow down your activity** d) Eat a lot

5. When your heart beats faster and you breathe harder during exercise, you are getting your...
a) Muscles stronger b) Bones heavier **c) Heart rate higher** d) Brain smarter

6. Push-ups help to strengthen your...
a) Legs **b) Arms and chest** c) Back d) Stomach

7. Squats help to strengthen your...
a) Arms b) Chest **c) Legs and buttocks** d) Back

8. Running helps to improve your...

- a) Balance b) Flexibility **c) Cardio (heart health)** d) Strength

9. Jumping rope helps to improve your...

- a) Flexibility b) Strength **c) Coordination** d) Balance

10. When you lift weights, you are doing...

- a) Strength training** b) Cardio exercises c) Flexibility exercises d) Balance exercises

11. You use your ___ to jump.

- a) Head **b) Legs** c) Arms d) Back

12. You use your ___ to throw a ball.

- a) Arm** b) Leg c) Head d) Stomach

13. You use your ___ to keep your balance.

- a) Arms b) Legs **c) Core (stomach and back)** d) Head

14. You use your ___ to hear instructions from your coach.

- a) Legs b) Arms **c) Ears** d) Stomach

15. You use your ___ to see the finish line in a race.

- a) Eyes** b) Legs c) Arms d) Stomach

16. It's important to wear proper shoes for exercise to avoid...
- a) Getting tired
 - b) Feeling hungry
 - c) **Injuries**
 - d) Feeling cold
17. What should you do if you feel dizzy or lightheaded during exercise?
- a) Keep exercising harder
 - b) Ignore it
 - c) **Stop exercising and rest**
 - d) Drink more juice
18. It's important to drink water before, during, and after exercise to stay...
- a) Warm
 - b) Strong
 - c) **Hydrated**
 - d) Cool
19. What should you do if you get hurt during exercise?
- a) Keep playing
 - b) Ignore it
 - c) **Tell your coach or parent**
 - d) Drink water
20. Which of these foods is a healthy choice for after exercise?
- a) Candy bar
 - b) Pizza
 - c) **Banana**
 - d) Soda
21. What is the opposite of "warm up"?
- a) Stretch
 - b) Strengthen
 - c) **Cool down**
 - d) Breathe
22. Which of these activities would help improve your flexibility?
- a) Running
 - b) Weight lifting
 - c) **Yoga**
 - d) Jumping jacks
23. A balanced meal after exercise should include:

- a) Mostly sweets
- b) Only vegetables
- c) A mix of protein, carbs, and healthy fats**
- d) Just a lot of water

24. What is the main benefit of getting enough sleep after exercise?

- a) More energy for games
- b) Helps your muscles recover**
- c) Makes you feel hungrier
- d) Improves your mood only

25. If you forget your water bottle at practice, what should you do?

- a) Keep exercising without water
- b) Borrow someone else's juice
- c) Ask your coach for water**
- d) Go home early

26. When lifting weights, it's important to use a weight that is: a) Too heavy so you get stronger faster

- b) So light you barely feel it
- c) Challenging but allows good form**
- d) The same weight as your friend

27. What is a good way to cool down after exercise? a) Stretching and walking

- b) Both A and C (Stretching and walking)**
- c) Jumping jacks and sprints
- d) Sitting down immediately

28. What does "cardio" refer to in exercise? a) Strength training

- b) Exercises that get your heart rate up**
- c) Balance exercises
- d) Flexibility exercises

29. Regular exercise can help you: a) Get better grades only
 b) Feel more tired
c) Feel more energetic and focused
 d) Make fewer friends
30. What is the best way to listen to your body during exercise?
 a) Ignore any aches or pains
 b) Push yourself as hard as you can
c) Pay attention to how you feel and stop if you feel pain
 d) Ask your friend how they feel and copy them

Appendix B: AMS

Statement	0 (Does not correspond at all)	1	2	3	4	5 (correspond exactly)
1. I don't know; I can't understand what I am doing in school.						
2. I can't see why I go to school and, frankly, I couldn't care less.						
3. Honestly, I don't know; I really feel that I am wasting my time in school.						
4. I don't feel like I have any reason to go to school.						
5. I go to school because my parents expect me to.						
6. I go to school because I want to please my parents.						
7. I go to school because I am afraid of getting into trouble if I don't.						
8. I go to school because I am afraid of disappointing my parents.						
9. I go to school because I want to prove to myself that I am capable of completing my high-school degree.						
10. I go to school because I want to show myself that I am an intelligent person.						
11. I go to school because I want						

to show myself that I can succeed in my studies.						
12. I go to school because I want to prove to myself that I am capable of doing well in school.						
13. I go to school because I believe that a high-school education will help me better prepare for the career I have chosen.						
14. I go to school because I think that a high-school education will improve my competence as a worker.						
15. I go to school because I believe that my high school education will enable me to enter the job market in a field that I like.						
16. I go to school because I believe that my high-school education will help me better prepare for the career I have chosen.						
17. Because I experience pleasure and satisfaction while learning new things.						
18. For the pleasure I experience when I discover new things never seen before.						
19. For the pleasure that I experience in broadening my knowledge about subjects which appeal to me.						
20. Because my studies allow me to continue to learn about many things that interest me.						

Appendix C: SEI

Statement	1 (Strongly agree)	2 (Agree)	3 (Disagree)	4 (Strongly disagree)
1. Overall, adults at my school treat students fairly.				
2. Adults at my school listen to the students.				
3. At my school, teachers care about students.				
4. My teachers are there for me when I need them.				
5. The school rules are fair.				
6. Overall, my teachers are open and honest with me.				
7. I enjoy talking to the teachers here.				

8. I feel safe at school.				
9. Most teachers at my school are interested in me as a person, not just as a student.				
10. The tests in my classes do a good job of measuring what I'm able to do.				
11. Most of what is important to know you learn in school.				
12. The grades in my classes do a good job of measuring what I'm able to do.				
13. What I'm learning in my classes will be important in my future.				
14. After finishing my schoolwork I check it over to see if it's correct.				
15. When I do schoolwork I check to see whether I understand what I'm doing.				
16. Learning is fun because I get better at something.				
17. When I do well in school it's because I work hard.				
18. I feel like I have a say about what happens to me at school.				
19. Other students at school care about me.				
20. Students at my school are there for me when I need them.				
21. Other students here like me the way I am.				
22. I enjoy talking to the students here.				
23. Students here respect what I have to say.				
24. I have some friends at school.				
25. I plan to continue my education following high school.				
26. Going to school after high school is important.				
27. School is important for achieving my future goals.				
28. My education will create many future opportunities for me.				
29. I am hopeful about my future.				
30. My family/guardian(s) are there for me when I need them.				
31. When I have problems at school my family/guardian(s) are willing to help me.				
32. When something good happens at school, my family/guardian(s) want to know about it.				
33. My family/guardian(s) want me to keep trying when things are tough at school.				
34. I'll learn, but only if my family/guardian(s) give me a reward. (Reversed)				
35. I'll learn, but only if the teacher gives me a reward. (Reversed)				

Text Readability of Machine Translation vs. Human Translation in Literary Texts: the Case for the Old Man and the Sea

Samad Mirza Suzani 

Assistant Professor, Department of English, Marvdasht Branch, Islamic Azad University, Marvdasht, Iran.
smirzasuzani@yahoo.com

Abstract

Comparing human and machine abilities for translating English literary texts into Persian and paying attention to the differences in the readability levels of the items to be translated into Persian by machine and human translators were the main impetus to conduct this study. Accordingly, the present study aimed to examine the legibility of two versions of Persian translation of the book “The Old Man and the Sea” (1951) by Google translate and human translator (namely, Azima, 2006). The comparison was done by Flesch Kincaid Grade, with different sub-criteria for evaluation. The findings suggest that both machine and human translations are very different based on Flesch Kincaid Grade criteria so that human translation may even be more readable. The results of this study can affect curriculum planning, especially in the field of translation. Another benefit of this study is that students could become familiar with their strengths and weaknesses in translation and hence it can be effective in evaluating translation students as would-be translators.

Keywords: Google Translate, Human translation, Machine translation, The Old Man and the Sea, Text readability

Citation: Mirza Suzani, S. (2024). *Applications of Language Studies*, 2(2), 202-225.

1. Introduction

In order to achieve good translation quality, a translator must pay attention to many aspects of the source language (hereafter SL) text that affect the form of the target language (hereafter TL) text. As McDonald (2020) asserts, “the purpose of translation is not to create a new work or a new writing, but to create a bridge between the author of the source language text and the readers of the target text” (p. 36). In other words, a translator does not translate the text into a new article, but must be able to act as a facilitator between the SL text and the TL text to convey the message.

Furthermore, McDonald (2020) argues that when it comes to the issue of translation evaluation, three factors need to be considered: acceptability, accuracy, and readability. Acceptability refers to “the fluency and naturalness of the translated text in accordance with the rules, language norms of the readers of the target text” (p. 23); accuracy means that “the transmission of the message or information of the source language is done accurately and honestly in accordance with the author's source language (p. 30), and readability is “the degree of ease of writing and understanding its meaning” (p. 30).

Based on McDonald (2020), the term acceptability gives the idea that the text must be accepted and understood by the reader of the TL, and that “if the cultures and target languages follow adherence to the norms and cultures of the source language, translation will be very appropriate” (p. 30). On the other hand, being accurate means “accurate transmission of meaning” (p. 33), while a translated text has a high level of readability if “it is easy to read and the reader can receive the text message of the target language without matching the text of the source language” (p. 33).

According to Crosbie (2013), the reader's understanding of the written text can be called readability of the text. Thus, the readability of a text is closely related to its content (vocabulary, complexity, and syntax), how it is presented (font size, line height, character spacing, and line length) and the ability to distinguish letters or characters from each other.

This study draws its significance from four major points: first, knowing translation software and their applications as an important point in translation; second, obtaining a better

evaluation of human translations, especially in relation to translators and interpreters; third, assessing readers' understanding of the translated texts; and finally, comparing the readability of text in human and machine translation with the help of readability software. Thus, the following research questions were raised in the present study:

1. How does machine translation differ from human translation in terms of rendering the readability of literary texts from English into Persian?
2. Is there any significant difference between the levels of item readability translated in Persian by machine and human?

2. Background

Some recent and clear studies have been conducted on the importance of readability. Rezaee and Norouzi (2011) conducted a study on readability and obtained a very high correlation between readability and performance indicators of readers. In another study on the relationship between the readability criteria and the comprehensibility level of source texts and their translations, Mujiyanto (2016) carefully studied the comprehensibility and readability of English texts and their translations and concluded that the level of comprehension of texts is attributed to their readers while the ability to read more depends on the text (cited in Acar & İşısağ, 2017).

It was also revealed that the comprehensibility and readability are intertwined so that they are sometimes used interchangeably, yet there are obvious differences between them. Acar and İşısağ (2017) maintain that perception is a reader-centered approach that includes comprehension, general knowledge, readers' cognitive networks, and the like. Kolahi and Shirvani (2012) state that although readability is a prerequisite for comprehension, it cannot be concluded that all readable texts are comprehensible or vice versa.

2.1. Studies on Text Readability Checker and Readability Formulas

The concept of readability has been widely discussed and introduced decades ago, and as Mujiyanto (2016) states, this concept was used as if it was “the only important topic that could be considered as a starting point for writing a text” (p. 5). Readability is a complex concept that relates both to the properties of texts and to the people and skills of readers.

Dubay (2004) recognizes readability as what makes some texts easier to read than others. Moreover, Acar and İşisağ (2017) argue that readability has most often been sacrificed for the sake of conveying the meaning of terms.

According to Dubay (2004), principles of readability include: using short, simple, familiar words; avoiding jargon; using culture-and-gender-neutral language; using correct grammar, punctuation, and spelling; using simple sentences, active voice, and present tense; beginning instructions in the imperative mode by starting sentences with an action verb, and using simple graphic elements such as bulleted lists and numbered steps to make information visually accessible.

Chall (1958) defines four elements for which the criterion of readability is important: vocabulary load, sentence structure, idea density, and human interest. Mühlenbock and Kokkinakis (2009) consider the criteria of readability as a combination that can include different aspects of readability. Vocabulary can be measured based on the word frequency ratio or the ratio or number of long words. In this vein, Moses was used to teach the translation model (Cohen et al., 2007) and SRILM to teach the language model (Stolcke, 2002).

Regarding the text readability, Crosbie (2013) states that higher readability plays a very important role in reading speed for all readers, especially those with low comprehension. Thus, increasing the readability level from average to good in readers with moderate or poor understanding can be very decisive in the success or failure of achieving goals. Besides, Crosbie (2013) adds that readability exists in both natural and programming languages in various forms. In programming, things like programmer comments, loop structure selection, and name selection can determine how easy it is for humans to read a computer program code.

Checking the readability of texts is important for many corporate and individual users, and as Brück et al. (2008) maintain, formulas for approximate text readability have a long tradition. According to Brück et al. (2008), the text readability checker is used to identify texts that are difficult to read and it can help writers write easy-to-read texts. In addition, it

often shows a universal readability score derived from a readability formula that describes the readability of the text numerically.

According to Klare et al. (1954), the empirical evidence obtained on the trap of the readability formulas indicates that a number of experts have been asked to re-examine the reliability of these formulas as a tool to measure people's understanding of specific texts. Meanwhile, Stephens (2000), states that readability tests can only measure text surface features so that qualitative factors such as difficulties in comprehension of words, composition, sentence patterns, property, abstractness, ambiguity as well as incoherence cannot be measured mathematically.

Maria et al. (2014) argue that most traditional readability approaches examine the surface properties of a text to determine the complexity of a text. Also, she believes that these readability criteria are based on the assumption that it relates superficial features to linguistic factors influencing language. She maintains that the average number of characters or syllables per word, the average number of words in a sentence, and the percentage of words that are not among the most common words in a language are all related to the words, syntax, and semantic complexity of the text, respectively.

Mujiyanto (2016) asserts that most of the research studies have been done on the subject from perspectives such as goal setting, psychological aspects as well as socio-cultural aspects. According to DuBay (2004), in the 1980s, there were approximately 200 readability formulas and more than a thousand studies on the implementation of formulas and the validity of their statistical theories. Among the major formulas, Flesch Reading-Ease test, first introduced in the 1950s, is still known as one of the most widely used items, the purpose of which is to reveal the level of ease of reading certain texts. The formula uses the average sentence length and the number of syllables per word to determine legibility; in addition, higher scores indicate greater ease of reading while lower scores indicate that passages are more difficult (Mujiyanto, 2016).

In addition to tests, a number of tools have been used to determine the readability level of texts which include the Flesch-Kincaid Grade Level with four levels including: Gunning-Fog score, Coleman-Liau index, Simple Measurement of Gobbledygook index

(SMOG), and Automatic Readability Index (ARI). These specific criteria are based on the average number of syllables and the number of words per sentence (Mujiyanto, 2016). The Gunning-Fog score, created by Robert Gunning in 1952, estimates the years of formal education required to understand a text on first reading. However, the SMOG index- founded by McLaughlin in 1969 for the same purpose, is considered more accurate than other formulas. Depending on the factor that the characters in each word are calculated faster, ARI is basically made to produce an approximate display of US level requirements to understand a text (Mujiyanto, 2016). Still again, the Coleman-Liau index, designed by Coleman and Liau (1975) as in previous formulas, relies more on letters calculated by computer programs than syllables (Maksymski, Gutermuth, & Hansen-Schirra, 2015).

A number of studies have explored the benefits of these measures and indicators. For example, Evanciew and Jones (1996) evaluated several textbooks used in high school and advanced technology programs related to legibility including grades (baseline equivalents), human interest, and writing style. Kolahi (2012) used the Gunning-Fog index to measure the level of readability to show that Persian translations of English textbooks are less readable than English textbooks. Meanwhile, Wolfer (2015) stated that the main purpose of “readability studies was to develop formulas that could be used to directly measure the readability of a text using text surface properties such as the average length of words or sentences” (pp. 36-37).

2.2. Studies on Readability in Machine Translation

Regarding readability in machine translation, readability and simplification of the text have been extensively studied in the field of computational tools and several criteria and approaches have been proposed in the literature. According to Koehn et al. (2003), common readability criteria use universal text attributes such as ratio, type/symbol, lexical consistency, and ratio of long versus short words as indicators. Global features, such as those listed above, generally require new approaches to the SMT decryption problem. In the model implemented by Docent, SMT is based on the phrase (Koehn et al., 2003). Decryption uses a local search approach that involves the complete translation of a complete document at any time. The initial state is improved by applying the summit operation strategy to find the maximum

(local) score function (Koehn et al., 2003). The three operations used are changing the translation of phrases, changing the position of two phrases, and redistributing phrases.

There are also a number of studies which use SMT techniques for monolingual translation (e.g., Ganitkevitch et al., 2011) and sentence compression (e.g., Knight & Marcu, 2000; Specia, 2010). In addition, there is a wide range of publications that simplify monolingual sentences using other methods for compressing and writing sentences (e.g., Daelemans et al., 2004; Cohn & Lapata, 2009). Regarding the integration of extensive contextual features in machine translation such as lexical consistency, Carpuat (2009) and Carpuat & Simard (2012) investigated the effect of lexical consistency on translation.

The use of word ambiguity in SMT is another common example of including textual information in the source section (Carpuat & Wu, 2007; Chan et al., 2007). Readability has also been measured as a text summary effect by automated user criteria studies (Margarido et al., 2008) and automated benchmarks (Smith & Jönsson, 2011). These studies indicated that readability was generally better in abridged texts than in original texts.

In another study, Genzel et al. (2010) studied the translation of poetry, the composition of news text, and the maintenance of the form of poetry in the translation of poetry in which the researchers employed decoding features such as rhyme and meters and introduced restrictions on the output of the target language in order to adapt to specific task features. Also, Tiedemann (2010) suggested stored models to successfully create consistent translations on domain compliance.

Previous work in machine translation and NLP has focused on evaluating the readability and simplification of the text. Jones et al. (1997) examined the readability of MT and ASR system output humanely. As for the text simplification, Hardmeier et al. (2013) explained a document-level decoder for SMT and mentioned a case study that utilized document-wide features to improve the readability of text. Stymne et al. (2012) introduces document-level features such as type/token ratios and lexical consistency as input to the MT system. Contrary to Stymne et al. (2012), Xu et al. (2016) designed a new training objective for SMT text simplification.

In his study, Teich (2003) found that the German translation is more passive than the original German texts, which is probably due to the influence of the original English. A similar event was observed in this study for MT readability in outputs; MT tended to show variance based on ST, indicating that there was a wave of ambiguity for the same word.

Brück et al. (2007) recognized legibility and comprehensibility as two main components that compare the quality of source texts as well as texts in technical and scientific translations. In order to evaluate the quality of the formula, Ateşman Reading Ease (1997) was used in technical and scientific translation. The ease of reading formulas and indicators used in both studies showed that the desired texts and translated texts are at the level of ease of reading. Easy-to-read results showed that technical and scientific translation and literary translation are a real challenge for translators.

Koponen et al. (2012) examined human perceived effort during the translation process. It was found that sentences with less effort, as shown in higher manual scores or shorter editing time, tended to edit more in terms of word form and simply replace words from the same word group, while sentences with lower scores tended to longer edits and take longer in terms of word order, and correction of mistranslated terms (Koponen et al., 2012).

For the time being, there are different types of machine translation including FastDic, Microsoft Translation, Babylon, and Google Translate. However, it appears that the readability of text has not been considered a lot in their translations. In other words, text readability has scarcely been evaluated in machine translation. With the advancement of machine translation technology, the comparison of human functions is often generalized and exaggerated. Therefore, it is important to be able to calculate the differences correctly. Indeed, meagerness of a study to investigate readability criteria to the level of comprehensibility of source texts and their translations has triggered carrying out this research study. Thus, the main aim is to explore whether the ease of reading and the variables of the level of English texts and their reading are related to their kinds of translation, namely human and machine translation. In this vein, the present study's major interest lays in the readability analysis of translated text from English into Persian considering HT and MT. In other words, the main aim of this study is to compare the ability of HT and MT to translate

English to Persian literary texts and to explore the differences between the levels of item readability in Persian translation performed by HT and MT.

3. Methodology

3.1. Research Design

The present study was comparative-descriptive in nature. In this vein, the closest translation to the original text was evaluated according to the readability of the text. This was done using automatic readability checker comprising formulas like Flesch Kincaid Grade.

3.2. Corpus

The corpus of this study included the celebrated English book “*The Old Man and the Sea*” by Ernest Hemingway (1952) and its Persian translations by Nazi Azima (2009) and Google translate. In this study, two translations of the first twenty pages of the above-mentioned English book by Azima and Google Translate were considered. To make a comparison, after providing the translated version by Azima, Google Translate web-based translation service was consulted for rendition of the same parts. Next, both translations were compared for the level of readability by Flesch Kincaid.

3.3. Procedures

The first twenty pages of the novel “The Old Man and the Sea” in English as well as its Persian translation by Nazi Azaima were thoroughly read to be analyzed critically. Then, the same pages were fed to Google Translate web-based translation service for translation to compare the readability of both translated texts. Eventually, Flesch-Kincaid Grade was used to calculate the readability of the Persian texts, as Flesch scores can indicate the readability of a text. The Flesch Reading Ease score is between 1 and 100, and the Flesch Kincaid score level represents the US education system. Also, the Flesch Reading Ease score was used to calculate the English text.

It should be noted that the ease/difficulty of reading is assessed by the reader's mental criteria. Moreover, the assessment of ease of reading is different in each language. For example, Flesch Ease Reading was used in this study for Persian; however, more than one

formula is used in English for assessing the ease of reading, (which in this study was Flesch-Kincaid). Flesch Kincaid Grade Level is a broad readability formula that assesses the approximate level of text readability. Previously, the Flesch Reading Ease score had to be converted to become a reading level. A modified version was used in the 1970s for ease of use. The Navy used it for its technical guidance used in training. It is now used for a much wider variety of applications. If a text has level 8 Flesch Kincaid, it means that the reader needs level 8 or higher to understand it. Even if they are an advanced reader, they have less time to read the content.

Considering that the readers' mental criterion is one of the most important factors in assessing the ease of reading, the ease/difficulty of reading a text is different for each individual reader. In this study, the dependent samples t-test statistical procedure was used to measure different levels of readability. As it was expected that the results of the t-test procedure could indicate different levels of readability, these results were explained inferentially. The Wilcoxon test was used for this inferential explanation to test the location of samples and to compare the locations of two populations using method samples.

4. Results

4.1. Addressing the First Research Question

In order to provide response to the first research question, Flesch Reading Ease Formula was used to compare the readability of English texts with machine translation and human translation. The readability ease of English texts was measured using Flesch Reading Ease Formula and the readability of Persian texts was measured using Flesch-Kincaid Grade Level. The obtained results are described in detail in Table 1.

According to Table 2, there is almost full equivalence between the source texts and target texts as far as the reading ease is concerned (90 to 100: very easy to read). Also, it can be inferred from the table that Flesch-Kincaid Reading Ease formula can be applied to measure the ease level of the target texts in Persian.

Table 1

Comparison of Reading Degree of Ease of Levels of Source Texts (ST) (English) and Targets Texts (Persian MT versus HT)

Criterion	Text	Language	Level
Flesch Reading Ease	ST	English	17.70
Flesch-Kincaid Grade Level	MT	Persian	20.92
Flesch-Kincaid Grade Level	HT	Persian	25.67

Table 2

Comparison of Flesch Reading Ease Scores and Flesch-Kincaid Scores

Interpretation of Flesch Reading Ease Scores	Interpretation of Flesch-Kincaid Scores
100.00-90.00: Very easy to read	90 – 100: Very easy to read
90.0-80.0: Easy to read	70 – 89: Easy to read
80.0-70.0: Fairly easy to read	50 – 69: Fairly difficult to read
70.0-60.0: Easily understood	30 – 49: Difficult to read
60.0-50.0: Fairly difficult to read	1 – 29 : Very difficult to read
50.0-30.0: Difficult to read	
30.0-0.0: Very difficult to read	

A comparison of Flesch Reading Ease Formula and Flesch-Kincaid Reading Ease Formula was made using Wilcoxon Test to compare the results inferentially. The results are presented in Table 3.

Table 3

Comparison of Flesch Reading Ease Formula and Flesch-Kincaid Reading Ease Formula (Wilcoxon Test)

	Mean	Std. Deviation	Z	P
Flesch Reading Ease Formula	18.31	17.982		
Flesch-Kincaid Reading Ease Formula	19.10	14.547		0.021

Based on the results of Wilcoxon Test in Table 3, it was confirmed that as far as they measure reading ease of the literary translation, Flesch Reading Ease Formula and Flesch-Kincaid Reading Ease Formula are compatible with each other (p = 0.021).

4.2. Addressing the Second Research Question

In order to address the second research question, Independent Samples t-test was used to measure the comprehensibility levels of the MT and HT for independent samples. The comparison of the readability levels of the MT and HT are presented in Table 4.

Table 4

Comparison of the Readability Levels of the MT and HT

Readability Level	N	Mean	Std. Deviation	t	P
MT	101	40.13	24.15		
HT	101	29.25	22.37	-4.571	0.001*

According to the results of Table 4.4, there is a significant difference between the readability levels of the MT and HT ($p < 0.05$). Thus, it can be concluded that the comprehensibility levels of HT is significantly lower than MT.

5. Discussion and Conclusion

Regarding the first research question, the results show that despite slight differences in the readability levels of both software, they are good references to compare the differences between SL and TL. Also, the results of Wilcoxon test confirm the compatibility of these two measures with each other. Moreover, the reading ease levels of texts decline as we move from source text in English to MT and HT in Persian, respectively. This can reveal the quality of long prose that the human translators tend to apply in order to preserve the language art and transfer the artistic tone and impact on the reader, while MT might solely be confined to transferring the meaning, using shorter sentences and more simple words.

The results related to the second research question reconfirm the same process with regard to the readability in Persian MT and HT, even though HT demonstrated more perplexity in comparison with MT. Accordingly, it seems that HT tends to act more independently in translation of ST. On the other hand, MT seems to be more dependent on the ST and does not modify it based on different relevant factors such as culture, intended meaning, etc. Moreover, this result might also originate from a lack of critical attitude in MT

towards the text and its translation process. In this vein, special techniques are required for these tasks to train prospective translation machines. Hence, it seems that terminological variations are regarded more meticulously concomitant to the process of critical thinking of the translator in HT, whereas this step is highly ignored in MT.

Brück et al. (2007) compared the readability and comprehensibility as two main components that clarify the quality of source texts as well as target texts to assess the quality of Ateşman Reading Ease formula in technical and scientific translation. Although they have not followed a similar path, the results of this study can be compared with their study in that in both studies, the Reading Ease formulas proved to be effective and accurate in evaluating the equivalence between the source texts and target texts. The Reading Ease and indices formulas employed in both studies revealed that both the target texts and the translated texts are over the general reading ease levels. The reading ease results indicated that technical, scientific, and literary texts translation poses a real challenge to the translators.

In relation to the degree of editing, Koponen (2012) and Koponen et al. (2012) examined human perceived effort during the process of translation. She understood that sentences engaging less effort, as demonstrated by higher manual grades or shorter editing times, were inclined to engage more edits in relation to word forms and simple replacements of words of the same word group, while sentences with low grades or long editing times engaged more edits in relation to word order, edits where the word group was altered, and modifications of mistranslated idioms (Koponen, 2012; Koponen et al., 2012). The results of the current research study are in line with Koponen's (2012) findings in that the same problem of micro level and macro level editing was encountered in her study; hence, these factors should be taken into consideration more cautiously. However, the focus of this study was mainly on the readability of MT output and its comparison with HT, while Koponen (2012) studied the concept of human editing more deeply.

Schulz, Bernhardt-Melischinig, Kreuzthaler, Daumke, and Boeker (2013) compared three different kinds of SNOMED CT translations from English to German with the use of specialized medical translators; a free web-based machine translation service; and medical students. Average ratings of linguistic accuracy did not vary meaningfully between HT

situations. Their results are consistent with the current study. In comparing MT to HT, the linguistic accuracy differed in support of the HT and concerning content fidelity, it differed similarly in support of the HT.

In a similar vein, Yamada (2014) assessed the quality of Google Statistical Machine Translation (SMT) by examining college language students' post-editing (PE) performance. Based on his findings, learners missed about 7 errors uncorrected in their final outputs. The data also revealed only a loose correlation between the learners' general translation ability and their post-editing performance. While learners who had weak marks in a traditional translation course were indicated to be unqualified post-editors, learners who achieved good marks were not always qualified post-editors either (Yamada, 2014). In the same vein, the existence of errors in MT is highlighted in the present research study. Accordingly, the proposed notion remains the same in both studies in that human translators have an important role in the translation process that should not be taken for granted and this accuracy of HT is better to be transferred to educational systems for training specialized translator machines.

Daems et al. (2015) conducted a post-editing study for general text types from English into Dutch accomplished by masters' students of translation. They recognized six various post-editing effort indicators and understood that various types of MT errors anticipated various post-editing effort indicators. However, they investigated different factors compared to the current study, a notion is similar in both studies; the highlighted matter is remained errors as an influence of MT products. Therefore, readers and translators shall recognize all potential errors in MT outputs in order to yield more qualified products.

In another related study, Comparin and Mendes (2017) revealed the findings of a research containing an error annotation task of a body of machine translations from English into Italian. They compared error types discovered in raw MT and post-edited texts. They also recognized frequent and critical errors and perceived the errors' occurrence at different phases of the translation procedure. One relevant result was that 85% of the errors discovered in the raw MT were correctly edited through human post-editing; also, fluency errors reduced, but a moderately high number of accuracy errors were not revised. In comparison with the results of the current study, their findings denoted the influence of MT on PE

products, highlighting the unrevised errors that requires a more critical attitude of the human translators. In a similar vein, Ahrenberg (2017) highlighted the limitations of MT outputs in the context of English and Swedish languages and claimed that PE products are not even comparable with HT in respect of quality.

Acar and İsisığ (2017) conducted a study on the readability and comprehensibility levels of technical and scientific texts in English and their Turkish translations using Flesch Reading Ease, Gunning Fog, Flesch-Kincaid Grade Level, the Coleman-Liau index, the SMOG index, Automated Readability index, and Linsear Write Formula. They compared and contrasted two components (i.e., the readability and comprehensibility levels of technical and scientific texts) in the target language and the source text. They applied Atesman Reading Ease Formula to estimate the reading ease of the translated texts in Turkish and they measured the comprehensibility levels of the source texts and target texts by using a checklist, including source texts and corresponding questions that were directed to 43 English teachers. One text was translated using Google translation. Their results can be compared with the present study in using Flesch Reading Formula which was found compatible to a Turkish Reading Ease Formula. In the same vein, this formula is found to be compatible with Persian Reading Ease Formula. However, they compared the comprehensibility levels of the source texts and the target texts that were found to be higher than the readability of the texts. The comprehensibility of the target texts was also discovered to be higher than that of the source texts. Moreover, a statistical difference was met between the readability and comprehensibility levels of the texts. Google translate had the lowest comprehensibility level.

Čulo, Hansen-Schirra, and Nitzke (2017) carried out a research study in which they contrasted post-editing and human translation with regard to the aspect of term translation within the field of LSP. They used the perplexity coefficient to analyze terminological variation in term translation from English to German. Relating to the achieved results of the present study, this requires a critical approach toward the outcomes of MT. It also tackles translators' community to have a more cautious use of the updated technologies.

Based on the results of the current study, there are different levels of ease of reading in each text, which shows the lowest to highest degree of ease in the source text, MT text and

HT text, respectively. Moreover, according to the obtained results, there is almost a perfect equivalent between the source and target texts as far as ease of reading is concerned (90 to 100: very easy to read). It can be inferred that the Flesch-Kincaid Reading Ease formula can be used to measure the ease level of the target texts in Persian.

To measure different levels of readability between machine and human translations, Independent Samples t-test was run and the results showed that there is a significant difference between MT and HT readability levels ($p < 0.05$). In this regard, the level of HT perception is significantly lower than MT. Of course, in this case only one language was examined in two translation models, which could be considered a potential downside of the current study. Another potential drawback of this study is the small sample of the study. The generalizability of the evaluations can be tackled by this drawback. Also, the assessment method was mainly based on the readability methods and did not consider other modes and dimensions of translation which could be compared and contrasted. Hence, the results are highly dependent on the method through which MT products were provided solely.

Assessing the readability of the text can affect the quality of text context. This means that the use of specific words and destination grammars, especially in the form of poetry, indicate the level of knowledge of student and the quality of the educational system. So it can be concluded that this study can affect curriculum planning, especially in the field of translation. This research study can also be effective in evaluating students in the field of translation as a translator. This means that another benefit of this research is the familiarity of students with their strengths and weaknesses in translation.

Along with the results of this study, an important question will be asked in the field of empirical corpus-based translation research. The unsolved issue is whether all types of texts tend to exhibit similar MT results concerning the linguistic properties or they may differ based on the subject matter. Therefore, a staunch need appears to further shift into translation studies in order to attain a comprehensive insight of the potential potpourri between MT and HT. Some modes of MT might be more compatible with target language; while, some other modes may display a contrary outcome. However, this notion cannot be judged with this small-scale study and requires deeper investigations.

In line with the above-mentioned results, another issue that should be taken into consideration is taking a step toward the progression of online translation tools in order to produce more qualified outputs and to ease the translation process. If MT proliferates rapidly as a prevalent method in the translation world and demonstrate similar texts as ST in different procedures of translation, the produced features will be narrowed down to a constant reproduction of the same being in the long run. In other words, this event leads to a linguistic flatter. In this regard, the potential question is whether literary texts would undergo any changes considering the continuous exchange of MT used by the related readers. Due to the limitations in the generalizability of the results, the results of this study cannot be overstated and hence further generalization requires more extended and fuller in-depth research.

Conflict of interest

The author(s) certify/certifies that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in the present research paper.

References

- Acar, A., & İşısağ, K. U. (2017). Readability and comprehensibility in translation using reading ease and grade indices, *International Journal of Comparative Literature and Translation Studies*, 5(2), 47-53. DOI: <https://doi.org/10.7575/aiac.ijclts.v.5n.2p.47>.
- Ahrenberg, L. (2017). Comparing machine translation and human translation: A case study. Paper presented at *Proceedings of the Workshop on Human-Informed Translation and Interpreting Technology*, Varna, Bulgaria. DOI:10.26615/978-954-452-042-7_003
- Ateşman, E. (1997). Türkçe’de okunabilirliğin ölçülmesi. *A.Ü. Tömer Dil Dergisi*, 58, 171-174.
- Brück, T., Hartrumpf, S., & Hermann Helbig, H. (2008). A readability checker with supervised learning using deep syntactic and semantic indicators. In *11th International Multiconference: Information Society-IS*, pp. 92–97.
- Carpuat, M. (2009). One translation per discourse. In *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions (SEW-2009)*. Stroudsburg: Association for Computational Linguistics: 19-27. Retrieved February 20, 2023 from: <http://aclweb.org/anthology-new/W/W09/W09-2404.pdf>.
- Carpuat, M., & Simrad, M. (2012). The trouble with SMT consistency. Paper presented at *Proceedings of the 7th Workshop on Statistical Machine Translation*. Stroudsburg: Association for Computational Linguistics: 442-449. Retrieved February 20, 2023 from: <http://aclweb.org/anthology-new/W/W12/W12-3156.pdf>.
- Chall, J. S. (1958). *Readability: An appraisal of research and application*. Columbus: Bureau of Educational Research, Columbus, Ohio, USA.
- Cohn, T. and Lapata, M. (2009). Sentence compression as tree transduction. *Journal of Artificial Intelligence Research*, 34, 637–674. DOI:10.1613/jair.2655
- Coleman, M., & Liau, T. L. (1975). A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, 60(2), 283–284. <https://doi.org/10.1037/h0076540>

- Comparin, L., & Mendes, S. (2017). *Using error annotation to evaluate machine translation and human post-editing in a business environment*. Paper presented at the Proceedings of EAMT, Prague, Czech.
- Crosbie, T., French, T., & Conrad, M. (2013). Stylistic analysis using machine translation as a tool. *International Journal for Infonomics (IJI)*, 1(1). Retrieved June 25, 3023 from <http://uobrep.openrepository.com/uobrep/handle/10547/333145>.
- Čulo, O., Hansen-Schirra, S., & Nitzke, J. (2017). Contrasting terminological variation in post-editing and human translation of texts from the technical and medical domain. *Empirical Translation Studies: New Methodological and Theoretical Traditions*, 300(1), 183-206. DOI:10.1515/9783110459586-007.
- Daelemans, W., Höthker, A., and Sang, E. T. K. (2004). Automatic sentence simplification for subtitling in Dutch and English. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04)*, pages 1045–1048, Lisbon, Portugal.
- Daems, J., Vandepitte, S., Hartsuiker, R., & Macken, L. (2015). The impact of machine translation error types on post-editing effort indicators. Paper presented at *4th Workshop on Post-Editing Technology and Practice (WPTP4)*, Miami, Florida. <http://hdl.handle.net/1854/LU-6990271>.
- DuBay, W. H. (2004). *The principles of readability*. Costa Mesa, California: Impact Information. Retrieved August 12, 2022 from <http://www.impact-information.com/impactinfo/readability02.pdf>.
- Evanciew, C. E. P., & Jones, K. H. (1996). Using readability, human interest, and writing style to evaluate technology education textbooks. *Tech Trends*, 41(2), 37-38. <http://dx.doi.org/10.1007/BF02818816>.
- Ganitkevitch, J., Callison-Burch, C., Napoles, C., & Durme, B. V. (2011). Learning sentential paraphrases from bilingual parallel corpora for text-to-text generation. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 1168–1179, Edinburgh, Scotland. <https://aclanthology.org/D11-1108>.

- Genzel, D., Uszkoreit, J., & Och, F. (2010). "Poetic" statistical machine translation: Rhyme and meter. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp. 158–166, Cambridge, Massachusetts, USA. <https://aclanthology.org/D10-1016>.
- Hardmeier, C., Tiedemann, J., & Nivre, J. (2013). Latent anaphora resolution for cross-lingual pronoun prediction. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 380–391, Seattle, Washington, USA. Association for Computational Linguistics.
- Jones, K. H. (1997). Analysis of readability, interest level, and writing style of home economics textbooks: Implications for special need learners. *Journal of Vocational Home Economics Education*, 12(2), 13-24. Retrieved June 12, 2021 from <http://www.natefacs.org/Pages/v12no2/12-2-13%20Jones.pdf>.
- Klare, G. R., Nichols, W. H., & Shuford, E. H. (1957). The relationship of typographic arrangement to the learning of technical training material. *Journal of Applied Psychology*, 41(1), 41–45. <https://doi.org/10.1037/h0046945>.
- Knight, K., & Marcu, D. (2000). Statistics-based summarization — Step one: Sentence compression. In *National Conference on Artificial Intelligence (AAAI)*, pp. 703–710, USA, Austin, Texas.
- Koehn, P., Och, F. J., & Marcu, D. (2003). Statistical phrase-based translation. In *Proceedings of the 2003 Human Language Technology Conference of the NAACL*, pp. 48–54, Edmonton, Alberta, Canada. DOI:10.3115/1073445.1073462.
- Kolahi, S., & Shirvani, E. (2012). A comparative study of the readability of English textbooks of translation and their Persian translations. *International Journal of Linguistics*, 4(4), 344–361. <https://doi.org/10.5296/ijl.v4i4.2737>.
- Koponen, M. (2012). *Comparing human perceptions of post-editing effort with post-editing operations*. Paper presented at the 7th Workshop on Statistical Machine Translation, Montreal, Canada.

- Koponen, M., Aziz, W., Ramos, L., & Specia, L. (2012). Post-editing time as a measure of cognitive effort. Paper presented at *the Workshop on Post-editing Technology and Practice, San Diego, USA*.
- Maksymski, K., Gutermuth, S., & Hansen-Schirra, S. (Eds.). (2015). *Translation and comprehensibility*. Berlin: Frank & Timme Verlag für wissenschaftliche Literatur.
- Margarido, P., Pardo, T., Antonio, G., Fuentes, V., Aluísio, S., & Fortes, R. (2008). Automatic summarization for text simplification: Evaluating text understanding by poor readers. In *Anais do VI Workshop em Tecnologia da Informação e da Linguagem Humana*, pp. 310–315, Vila Velha, Brazil. DOI:10.1145/1809980.1810057.
- Maria A. C., & Dinu. L (2014). A Quantitative Insight into the Impact of Translation on Readability. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)*, pages 104–113, Gothenburg, Sweden. Association for Computational Linguistics. DOI:10.3115/v1/W14-1212.
- McDonald, S. V. (2020). Accuracy, readability, and acceptability in translation. *Applied Translation, 14*(2), 21–29. <https://doi.org/10.51708/apprans.v14n2.1238>.
- Mühlenbock, K., & Kokkinakis, S. J. (2009). LIX 68 revisited – an extended readability. In *Proceedings of the Corpus Linguistics Conference*, Liverpool, UK.
- Mujiyanto, Y. (2016). The comprehensibility of readable English texts and their back-translations. *International Journal of English Linguistics, 6*(2), 21. <https://doi.org/10.5539/ijel.v6n2p21>.
- Rezaee, A. A., & Norouzi, M. H. (2011). Readability formulas and cohesive markers in reading comprehension. *Theory and Practice in Language Studies, 1*(8), 1005–1010. <https://doi.org/10.4304/tpls.1.8.1005-1010>.
- Schulz, S., Bernhardt-Melischinig, J., Kreuzthaler, M., Daumke, P., & Boeker, M. (2013). sMachine vs. human translation of SNOMED CT terms. In *MEDINFO 2013 - Proceedings of the 14th World Congress on Medical and Health Informatics* (1-2

- Aufl., S. 581-584). (Studies in Health Technology and Informatics; Band 192, Nr. 1-2). <https://doi.org/10.3233/978-1-61499-289-9-581>.
- Smith, C., & Jönsson, A. (2011). Automatic summarization as means of simplifying texts, an evaluation for Swedish. In *Proceedings of the 18th Nordic Conference on Computational Linguistics* (NODALIDA'11), Riga, Latvia.
- Specia, L. (2010). Translating from complex to simplified sentences. In *Proceedings of the 9th International Conference on Computational Processing of the Portuguese Language*, 9th International Conference (PROPOR'10), pp. 30–39, Porto Alegre, Brazil.
- Stephens, C. (2000). All about Readability. Retrieved February 2, 2024 from <http://www.plainlanguage.com/newreadability>.
- Stolcke, A. (2002). SRILM – an extensible language modeling toolkit. In *Proceedings of the 7th International Conference on Spoken Language Processing*, pp. 901–904, Denver, Colorado, USA.
- Stymne, S., & Smith, C. (2012). On the interplay between readability, summarization and MTranslatability. In *Proceedings of the 4th Swedish Language Technology Conference*, pp. 71–72, Lund, Sweden.
- Teich, E. (2003). *Cross-linguistic variation in system and text: A methodology for the investigation of translations and comparable texts*. Berlin/New York: Mouton de Gruyter. DOI:10.1515/9783110896541.
- Tiedemann, J. (2010). Context adaptation in statistical machine translation using models with exponentially decaying cache. Paper presented at *Proceedings of the 2010 Workshop on Domain Adaptation for Natural Language Processing (DANLP)*. Stroudsburg: Association for Computational Linguistics: 8-15. Retrieved April 25, 2023 from: <http://aclweb.org/anthology-new/W/W10/W10-2602.pdf>.

Wolfer, S. (2015). Comprehension and comprehensibility. In K. Maksymski, S. Gutermuth, & S. Hansen-Schirra (Eds.), *Translation and comprehensibility* (pp. 33-52). Berlin: Frank & Timme Verlag für wissenschaftliche Literatur.

Yamada, M. (2014). Can college students be post-editors? An investigation into employing language learners in machine translation plus post-editing. *Machine Translation*, 29(1), 49–67. DOI:10.1007/s10590-014-9167-7.

Translation Quality Assessment of Persian Translations of Charles Dickens's *Great Expectations* Based on House's TQA Model

Nooshin Ahmadi Darani 

Master of Translation Studies, Payam-e-Noor University, Isfahan branch, Isfahan, Iran.
nsh.ahmadi1376@gmail.com

Abstract

This research paper examines two translations of Charles Dickens's *Great Expectations* by Yunesi and Akhavan, utilizing House's model. House's model of Translation Quality Assessment (TQA) provides a comprehensive framework for evaluating the quality of translations in terms of linguistic accuracy, cultural equivalence, and stylistic appropriateness. The selection of these two translations was based on their popularity and relevance to the study. Based on this model, overt or covert translation and the implication of cultural filters were determined. Moreover, the translation that effectively conveys the core meaning of the original text was also found. In this study, a comparative analysis of 50 pages of the translated texts compared to their original version was done. The main focus was on determining which translation was superior, more subtle, and more explicit. The study utilized House's model as a framework for evaluation, allowing for a structured and systematic approach for comparing the two translations. The findings revealed that Yunesi's translation was overt and superior because of fewer overt errors, and Akhavan's translation was covert. Cultural filters were applied in Akhavan's translated text in order to make proper translated texts for children's and teenagers' ages.

Keywords: Translation Quality Assessment, Covert translation, Overt translation

Citation: Ahmadi Darani, N. (2024). *Applications of Language Studies*, 2(2), 225-253.

1. Introduction

This study focuses on evaluating the quality of Persian translations of Charles Dickens's novel "*Great Expectations*" using House's Translation Quality Assessment (TQA) model. The study compares different Persian translations of the novel to analyze their accuracy and effectiveness in conveying the meaning of the original text. By examining the translations through a specific model, the article aims to provide insights into the strengths and weaknesses of each translation, offering a comprehensive overview of the topic and enhancing the reader's understanding of translation quality assessment in the context of literary works. This study was done based on a holistic and eclectic model of translation quality assessment that is proposed by House (1997).

This model analyses the original and translated text regarding the (a) Field of the text (subject matter and social action), (b) Tenor (author's provenance and stance and social role relationship), (c)-Mode (medium and participation), (d)-Genre and (e) the Function of the text. This model was revised based on Hallidayan systematic functional theory (1978). The researcher tried to make an appropriate ground for another translation assessment. The significance of evaluating translated texts lies in the need for translators to progress and for translated texts to be reviewed and enhanced. This issue has motivated the researcher to assess the chosen translated texts.

In the article comparing translations with the original text to determine the quality of translations, several key problems may arise. Firstly, the challenge of accurately capturing the nuances and cultural context of the original text in the translated version can be a significant hurdle. This can lead to potential misinterpretations or loss of meaning in the translated work.

Secondly, the implications of cultural filters in translations can also pose a problem. Different cultures may have varying perspectives, values, and linguistic nuances that can impact how a text is translated and understood. This can result in a loss of cultural authenticity or misrepresentation in the translated work.

Lastly, distinguishing between overt and covert translations can be a complex task. Overt translations aim to directly translate the words and phrases from the original text, while

covert translations focus on conveying the underlying meaning and intent of the text. Determining which approach is more suitable for a particular text can be challenging and may impact the overall quality of the translation.

As a researcher in the field of translation studies, conducting research on House's Model of Translation Quality Assessment in the context of evaluating translations of Charles Dickens's *Great Expectations* holds several significant implications. First of all, House's model is one of the most well-known and established models in the field of translation quality assessment. By utilizing this model, the research can benefit from an extensive theoretical foundation and practical application of TQA, ensuring credibility in the evaluation of the translated texts. So the researcher tried to design the study in order to apply a well-known translation quality model. Secondly, this study is done to understand the translation quality of Dickens's *Great Expectations*. Studying the translations can contribute to a deeper comprehension of the translation techniques used to capture its essence in different languages.

By utilizing House's model, one can provide a systematic analysis of the translations' fidelity, fluency, and other relevant dimensions, shedding light on how well the translations convey the original work's meaning, style, and impact. Then, this can help translators, scholars, and publishers understand the challenges. Thus, this study is done to identify the strengths and weaknesses in translation texts. Moreover, it is a translation quality assessment research study by applying House's model to evaluate the translations of literary works like *Great Expectations*. Finally, the researcher has found this subject to be helpful for assisting other translators, publishers, and researchers. Additionally, readers who rely on translated texts can make more informed decisions when selecting a translation of *Great Expectations* based on the analysis and evaluation provided in this paper. The research questions are as follows:

- 1) According to House's model of TQA, which selected translated texts of Dickens's *Great Expectations* excelled in conveying the essence of the original work?
- 2) According to House's model of TQA, are the selected translated texts overt or covert?

- 3) According to House's model of TQA, what is the implication of cultural filter in the selected translated texts?

2. Review of Literature

House's model is a well-known framework in the field of translation and comparative analysis used for evaluating and analyzing literary translations. This model was introduced by Julian House in 1977 and consists of four main dimensions: Meaning, Language, Culture, and Time.

1. *Meaning Dimension*: This dimension focuses on accurate and precise transfer of meaning from the original text to the translation. It includes conveying concepts, ideas, and the main messages of the text.
2. *Language Dimension*: In this dimension, attention is given to the structure and linguistic features of the original text, and efforts are made to preserve these features in the translation.
3. *Culture Dimension*: This dimension emphasizes the cultural and social influences on translation. The translator must pay attention to the cultural concepts and idioms of the original text and try to preserve these dimensions in the translation.
4. *Time Dimension*: This dimension looks at temporal and historical influences on translation. The translator must consider when and in what context the original text was written and how to incorporate these aspects into the translation.

For example, suppose you want to evaluate a translation of a famous poem. By using House's model, you can look at the meaning, linguistic structure, cultural idioms, and temporal aspects to offer a precise and comprehensive translation of the poem that considers all these dimensions. This model helps you deliver a high-quality and faithful translation that aligns with the original text.

Conducting research based on House's Model of translation quality assessment to evaluate translations of Dickens's *Great Expectations* holds great significance in terms of increasing knowledge in translation studies, contributing to the field's evaluation

methodologies, and assisting translators, publishers, and readers in acquiring a better understanding of the translation quality of this literary work. Because of the importance of this topic, numerous studies have been done in this field. Several prior studies have employed House's Model of Translation Quality Assessment to assess the translated texts. These studies have concentrated on evaluating the overt and covert translations and comprehending the impact of cultural filters. These articles have made valuable contributions to the field of translation studies. Here are some examples:

Mozaheb and Salmaki (2020) paid attention to the quality of Dickens's *Great Expectations* translated by Maryam Dastum by using House's model of TQA. The analysis was done by investigating lexical means, syntactic means, and textual means, in three parts of the field, tenor, mode, and genre. The researchers compared the source text with the target text again and investigated the field, tenor, mode, and genre. The findings showed that this translation was covert translation and the quality of translation was normal.

Tabrizi, Chalak, and Taherioun (2014) investigated Orwell's *Nineteen Eighty-Four* quality based on House's model of TQA (1997), which was translated by Balooch (2004) into Persian. In this research study, the source and target text profiles were investigated and the function of the texts was given. Then, the source text and target text were compared and the overt errors were analyzed. The substitution was the most and the breaches of TL systems were the lowest frequent of overt errors. Based on this paper, the translation could not be an overt translation because of some overt errors and cultural filters; rather, it was a covert translation.

Charkhtab and Monsefi (2019) investigated the quality of Wilhelm's Tyrone *Big Bad Bully* Persian translation which was done by Jaberfard (2016) according to House's model of TQA (1997). The researchers compared the source text with the target text and analyzed the mismatches. Moreover, the field, tenor, mode, and genre of the source and target texts were determined. It was concluded that addition was the most frequent and breaches of TL system were the least frequent instances of overt errors, in addition, the translation was covert because of several errors and cultural filters.

Another thesis was written in this area by Masoudnia (2017) which was an investigation of Persian translations of “The Girl on the Train”. The researcher concluded that Salary’s translation was more overt than Mosavi’s translation. Also mosavi’s translation was more comprehensible based on House’s model of TQA (1997) (Masoudnia, 2017).

There is another paper about the application of House’s model of TQA (1997) in Persian translations of Orwell’s *Nineteen Eighty-Four*. The profiles were compared together and the mismatches were assessed and omissions and substitutions were the most frequent errors. Also because of the existence of cultural filters, it was a covert translation (Heidari Tabrizi et al., 2013).

Another study, which applied the TQA model to evaluate the translation quality of cultural references in literary texts, was done by Kamala (Kamala, 2007). He examined how overt and covert translations handle cultural filters and the implications for the reception of these references by the target audience. The article provided significant insights into the challenges faced by translators in maintaining cultural richness while ensuring target reader comprehension.

Another article focused on assessing the quality of covert translations using House's TQA model. It made an attempt to investigate the effects of cultural filters on covert translations and the resulting impact on translation quality. The study contributes to our understanding of the complexities and challenges of conveying cultural nuances in translated texts (Wu, 2012).

Another study investigated the implications of cultural filters in Chinese-English translations by employing House's TQA model (Chen, 2014). The impact of cultural differences on translation quality, highlighting both the strengths and weaknesses of the translated texts was explored. It offers valuable insights into the decision-making processes and challenges faced by translators.

These articles demonstrate the relevance and effectiveness of House's TQA model in evaluating translation quality, particularly in the context of overt and covert translations and the implications of cultural filters. They contribute to the advancement of translation studies

by providing empirical evidence, practical recommendations, and a deeper understanding of the complexities involved in translating cultural elements. Additionally, identifying strengths and weaknesses of the translated texts was another important issue explored in these articles. These studies are definitely helpful for translators to obtain insights regarding the decision-making processes when dealing with overt and covert translations and cultural filters.

3. Method

3.1. Theoretical Framework

This study is done based on House's model of Translated Quality Assessment (1997). This model was proposed according to the theories of language use that focus on linguistic and extra linguistic aspects. According to House (1997), equivalence is important in this model and must be preserved as functional equivalence or as preserving the meaning. This model of TQA analyses the source and the translated texts by focusing on three aspects of texts including field, tenor, and mode that are gathered in the register segment. The genre segment was added to this model later. Moreover, the field segment includes subject matter and social action, the tenor segment includes the author's provenance and stance, social role relationship, and social attitude, and mode segment includes medium and participation. This study involves both qualitative and quantitative research design, while the data collection method based on the research objectives and nature of the study is observation and analysis of existing sources. The researcher used a systematic method in which 50 pages were selected, spanning from the first chapter to the last chapter of the text. The selection process involved choosing pages at regular intervals, with each page being 9 pages apart from the previous one. This method ensures a representative sample of the text while managing the scope of the analysis.

The researcher aims to ensure that the findings are valid and reliable. Measures must be taken to minimize bias, ensure the accuracy of data collection and analysis methods, and demonstrate the trustworthiness of the research findings. Thus, the researcher ran the Chi-Square test. The qualitative approach was used to assess the cultural filters in translation. In order to do this, the researcher analyzed the translation process through the lens of cultural differences and how they are reflected in the translated text. By examining how cultural

filters are applied and how they impact the overall quality of the translation, the researcher determined the effectiveness of the translators in conveying the intended meaning and cultural nuances of the original text.

3.2. Corpus

Great Expectations of Charles Dickens was selected as the source text published by Barnes and Noble (2004). Two target texts were translated by Yunesi (2020) and Akhavan (1996). The original text had 544 pages in the first edition in 1861, so the researcher decided to bind the sample size and tried to focus on 50 pages which were chosen systematically from the first chapter to the last chapter and the interval was 9 pages. In other words, the researcher selected the first page, in every 9 pages to have 50 pages as the research sample. The researcher used this sampling to enhance the validity of the study, because translators would make these overt errors in different parts of the translated text for some reasons like time pressure. Also, choosing the sample of study from the first pages of the translated text might not be credible because the translator is so motivated and precise in the beginning of the translation.

3.3. Procedures

This study was done to distinguish the application of House's model of TQA on Dickens's *Great Expectations* and its two Persian translated texts by Yunesi and Akhavan. In House's model of Translation Quality Assessment, the assessment of translated texts involves analyzing various aspects of the translation, such as field, tenor, mode, genre, and participation. The research procedure was proposed for assessing translated texts based on this model. These factors help evaluate the appropriateness of the translation in terms of subject matter, register, style, genre conventions, and the intended recipients.

The researcher conducted a comparative analysis between the source text and the translated texts. Then, she identified areas where changes have occurred, including omissions, additions, distortions, breaches of the target language (TL) system, or slight changes in meaning. These elements were assessed against the established criteria to

determine the quality of the translation. The researcher applied both quantitative and qualitative analysis techniques to assess the translation quality.

In the first step, the researcher determined both the source and the target text profiles. Next, the translated texts were compared with the original text separately and then both of the target texts were compared with the original text simultaneously. In addition, the cultural filter elements were determined. The data were analyzed to find answers to the research questions. The researcher chose two versions of the Persian translation of *Great Expectations* by Yunesi (Dickens, 2005) and by Akhavan (Akhavan, 1996).

4. Results

4.1. ST Analysis

These details were derived from the original text based on House's model which includes the elements of field, tenor, mode, genre, and participation to create a profile of the original text.

Table 1

The Details Derived from the Original Text based on House's Model

Individual Textual Function		Examples
Field	Lexical Means	Joe greatly augmented my curiosity. (p. 41)
	Syntactic Means	My sister was... (p. 50)
	Textual Means	May I venture to congratulate you (p. 176)
The field of " <i>Great Expectations</i> " revolves around themes of social class, ambition, love, and identity. The novel explores the journey of the protagonist, Pip, as he navigates through various societal expectations and personal desires.		
Tenor	Lexical Means	As we were going with our candle

		along the dark passage (p. 140)
	Syntactic Means	The second greatest surprise I have ever had in my life was seeing him... (p. 113)
	Textual Means	I thank you ten thousand times (p. 302)

The tenor of the novel focuses on the relationships between characters, particularly the complex dynamics between Pip, Estella, Miss Havisham, and Magwitch. Dickens delves into the emotional and psychological aspects of these relationships, highlighting the impact they have on the characters' lives.

Mode	Lexical Means	“Oh yes”, yes he returned (p. 212)
Medium: complex	Syntactic Means	“O! Don’t cut my throat, sir,” I pleaded in terror (p. 32)
	Textual Means	I was so humiliated... that tears started to my eyes (p. 86)

The mode of “*Great Expectations*” is predominantly narrative, as Dickens tells the story through the perspective of Pip. The novel is characterized by vivid descriptions, intricate plot twists, and rich character development, all of which contribute to its immersive storytelling.

Participation	Lexical Means	I asked him/ he answered
	Syntactic Means	“Tell us your name!” Said the man. “Quick!” (imperative) (p. 32).
	Textual Means	The four little white crockery poodles on the mantel shelf (p. 50)

Readers actively participate in the text by analyzing the characters’ motivations, societal critiques, and moral dilemmas presented in the novel. The reader's engagement with the text enhances their understanding of the themes and messages conveyed by Dickens.

Genre	This story is a bildungsroman and a drama, which it considers psychological and moral situations. This novel surrounds detailed life events of main role.
-------	---

“*Great Expectations*” falls under the genre of Victorian literature, specifically belonging to the genre of the Bildungsroman, or coming-of-age novel. It combines elements of social commentary, moral exploration, and psychological depth, making it a classic example of 19th-century English literature.

4.2. Source Text Profile

The researcher started by thoroughly examining the linguistic features of the source text, including its vocabulary, grammar, sentence structure, and overall coherence. The researcher paid close attention to Dickens's writing style, which often features complex sentences, rich vocabulary, and intricate narrative techniques.

Table 2

The Source Text Profile

Field:	Subject matter	Novel
	Social action	General & popular
Tenor:	Author's provenance and stance	Writer & novelist
	Social role relationship	Symmetric
	Social attitude	Informal
Mode:	Medium	Complex
	Participation	Complex
Genre:	Moral & psychological	
Function:	Interpersonal & ideational	

Field: *Great Expectations* belongs to the field of literature, specifically the genre of Victorian novels. It deals with themes of social class, ambition, love, and identity in 19th-century England.

Tenor: The relationships between characters in *Great Expectations* are complex and often driven by social expectations and personal aspirations. The protagonist, Pip, navigates interactions with characters such as Estella, Joe Gargery, and Miss Havisham, reflecting various aspects of human behavior and emotion.

Mode: The mode of *Great Expectations* is primarily narrative, with the story unfolding through a first-person perspective as Pip recounts his life story. Dickens employs

vivid descriptions, rich character development, and intricate plots to engage readers and convey the novel's themes effectively.

Genre: *Great Expectations* falls within the genre of Bildungsroman, a literary form that focuses on the moral and psychological growth of the protagonist from childhood to adulthood. It also encompasses elements of social critique and romance, adding layers of complexity to the narrative

Function: The function of *Great Expectations* goes beyond mere entertainment; it serves to explore the human condition, societal norms, and the pursuit of self-identity, and fulfillment. Through its characters and plot twists, the novel prompts readers to reflect on themes of ambition, morality, and the impact of one's past on his/her future.

In summary, Dickens's *Great Expectations* is a multifaceted work that operates within the field of literature, portraying intricate relationships through its dynamic characters and narrative mode. As a Bildungsroman, it serves the function of not only telling a compelling story but also delving into deeper themes and societal observations that continue to resonate with readers across generations.

4.3. Target Text Profile

The researcher assessed the degree of linguistic equivalence between the source text and its translations by Yunesi and Akhavan. The researcher scrutinized the translated texts to determine how effectively they capture the linguistic nuances and stylistic features of the original text.

Field: The field of "*Great Expectations*" by Dickens revolves around themes of social class, ambition, and identity. The translations by Yunesi and Akhavan maintain a similar field, focusing on these same themes which are adapted for their respective audiences.

Tenor: The tenor of Dickens's original work is likely formal and literary, while the translations by Yunesi and Akhavan adjust the tenor to suit the cultural and linguistic norms of their target audiences.

Mode: The mode of communication in “*Great Expectations*” is through written text, conveying complex narratives and character development. The translations by Yunesi and Akhavan would also be in written form but may be adapted to better resonate with their readers.

Genre: “*Great Expectations*” is a classic novel, falling under the genre of literary fiction. The translations by Yunesi and Akhavan also fall under the genre of literary fiction but may incorporate elements that cater to their specific audiences.

Function: The function of “*Great Expectations*” is to entertain, provoke thought, and explore societal issues. The translations by Yunesi and Akhavan maintain these functions while also serve to bridge cultural and linguistic gaps between the original work and their audiences.

Table 3

The Target Text Profile

Field:	Subject matter	Novel
	Social action	General & popular
Tenor:	Author's provenance and stance	Writer & translator
	Social role relationship	Symmetric
	Social attitude	Informal
Mode:	Medium	Complex
	Participation	Complex
Genre:	Moral & psychological	
Function:	Interpersonal & ideational	

Overall, while the core themes and messages of “*Great Expectations*” may remain consistent across the original work and its translations by Yunesi and Akhavan, the adaptations in field, tenor, mode, genre, and function will be made to ensure the works resonate with their respective audiences effectively.

4.4. Comparison between ST and TT

4.4.1. Elements which were not Translated in Akhavan’s Translation

These include idiomatic expressions, wordplay, or culturally specific references that may not have direct equivalents in the target language. Translators may opt for approximations or cultural substitutions, which could result in a loss of nuance or richness of the original text.

Table 4

Comparison between ST and TT

ST	TT
1- I pointed to where our village lay, on <u>the flat in-share among the alder trees and pollards</u> , a mile or more from church (p. 32).	من به طرف دهکده مان که حدود یک مایلی کلیسا واقع شده بود اشاره کردم (اخوان، ص. ۹).
2- What with my feelings and what with soup-suds I could at first see no stars from the chaise-cart (p. 77).	هیچ ستاره‌ای توی آسمون نمی دیدم (اخوان، ص. ۴۱).

As it is clear by comparing the source texts and translated texts, there are some untranslated and removed elements in Akhavan's translations. The researcher claimed that Akhavan's translated text was abbreviated to make a short story and that is why the target text is much shorter.

Table 5

Additions in Akhavan's Translation

ST	TT
1- "Now lookee here" said the man. "Where's your mother?" (p. 32).	از من پرسید که پدر و مادرم کجا هستند (اخوان، ص. ۱۰).
2- I believe they were fat... (p. 32).	اما جوجه‌هایی که داشتیم به نظر من چاق بودند (اخوان، ص. ۱۰).

4.4.2. Additions in Akhavan's Translation

Translators may sometimes add supplementary materials to provide context, clarify ambiguities, or enhance the reader's understanding. These additions could include footnotes, glosses, or explanatory passages. While intended to aid comprehension, excessive additions

might alter the original text's tone or disrupt its flow, impacting the reader's engagement with the translated work.

There are some added elements in Akhavan's translated text. The number of additions was less than removed items, in order to make the story localized and emphasize the fluency and ease of the target text not just translating word for word.

4.4.3. Distortions in Akhavan's Translation

Distortion occurs when the translator misinterprets or alters the meaning, tone, or intent of the source text. This can result from linguistic limitations, cultural biases, or subjective interpretations. Distortions may manifest as changes in characterization, plot details, or thematic emphasis, leading to a divergence from the author's original vision.

Table 6

Distortions in Akhavan's Translation

ST	TT
1- "Show <u>us</u> where you live" (p. 32).	جایی را که زندگی می کنی، به من نشان بده (اخوان، ص. ۹).
2- The <u>ringing and roaring</u> stopped (p. 59).	همه ی سروصداها از بین رفت (اخوان، ص. ۲۹).

Distortion refers to the alterations or modifications that occur during the process of translation, resulting in a deviation from the source text. It is an important concept in translation quality assessment, which aims to evaluate the fidelity and accuracy of translations. It means that the translator did not remove or add any elements but decided to change them. In Akhavan's translation, distortion occurred to bring fidelity, adequacy, and acceptability.

4.4.4. Breaches of TL System in Akhavan's Translation

Breaches of TL norms refer to violations of linguistic conventions, stylistic norms, or grammatical rules in the target language. These breaches can occur due to the errors in translation, awkward phrasing, or attempts to replicate source language structures that are

incompatible with the target language. Such breaches undermine readability and authenticity, detracting from the overall quality of the translation.

Table 7

Breaches of TL System in Akhavan's Translation

ST	TT
1- " <u>You young dog</u> ", said the man, licking his lips... (p. 32).	در حالیکه لب هایش را می لیسید گفت: « <u>ای سگ</u> جون... (اخوان، ص. ۹).
2- Evening mists were rising now and in all the board expanse of tranquil light they showed to me <u>I saw</u> no shadow of another parting from her (p.542).	در فضای آرام آن روز بعد از ظهر، احساس کردم که هیچ وقت از او جدا نخواهم شد (اخوان، ص. ۲۳۰).

Breaches of the target language system refer to instances where the translation deviates from the accepted norms and conventions of the target language. These breaches can negatively impact the quality and effectiveness of the translated text. In Akhavan's translation, there were some changed elements that were not generally used in the target language. The researcher considered these breaches were made to rise the comic atmosphere, but they affected the translation negatively.

4.4.5. Slight Changes in Meaning in Akhavan's Translation

Even with a meticulous attention to details, translations often entail minor shifts in meaning or nuances due to linguistic differences between the source and target languages. These changes may result from lexical choices, syntactic structures, or cultural adaptations which are necessary for effective communication in the target language. While some alterations are inevitable, translators strive to minimize deviations from the original text's intended meaning.

Slight changes in meaning refer to instances where the translated text deviates slightly from the original meaning of the source text. These changes may be unintended and can have an impact on the accuracy and fidelity of the translation. There were some slight changes in meaning in Akhavan's translation, these changes in meaning were not effective in the narrative.

Untranslated elements in Yunesi's translation were less than those in Akhavan's translation. As it is clear by comparing the source text and target text, the reason could be the translator intention to reduce the items which were ineffective in the narrative.

Table 8

Slight Changes in Meaning in Akhavan's Translation

ST	TT
1- "What fat cheeks you ha got"? (p. 32)	چند تا جوجه ی چاق دارید؟ (اخوان، ص. ۹)
2- I kissed her cheek as she turned it to me (p. 113).	در حالیکه اونو بوسیدم، او هم منو بوسید (اخوان، ص. ۶۸)

Table 9

Elements which were not Translated in Yunesi's Translation

ST	TT
1- Pip, sir" "Once more" said the man staring at me. "Give it mouth" "Pip, Pip, sir".	«پپ، آقا» ----- «به ما بگو کجا زندگی می کنی زود آنجا را نشان بده!» (یونسی، ص. ۱۷).
2- On a <u>memorable</u> raw afternoon towards evening... (p. 32).	در بعدازظهری سرد و مرطوب حوالی غروب بدست آمده باشد. (یونسی، ص. ۳۲).

Addition in Yunesi's translation was more than that in Akhavan's translation. The translator decided to make the content clearer based on comparing the source text and target text. The addition did not influence the story line.

Distortion occurred due to various factors such as linguistic, cultural, and situational differences between the source and target languages. Also, complete equivalence is often unattainable. In Yunesi's translation, the number of distortions was more than that in Akhavan's translation.

In House's TQA model, breaches of the TL system are considered as part of the acceptability level of assessment. The acceptability level focuses on the stylistic and cultural appropriateness of the translation in the target language. It aims to determine whether the translation adheres to the norms and conventions of the TL and if it is natural, fluent, and readable for the target audience.

Table 10

Additions in Yunesi's Translation

ST	TT
1- I believe it was settled you should meet me? (p. 266)	گمان میکنم قرار بر این شد که مرا در <u>چاپارخانه</u> ملاقات کنید؟ (یونسی، ص. ۳۴۴)
2- Under what altered circumstances I should next see those rooms, if ever (p. 428).	از <u>خود</u> پرسیدم: تو می گویی تحت چه شرایط متفاوتی این آپارتمان را دوباره خواهیم دید؟ (یونسی، ص. ۵۱۸).

Table 11

Distortions in Yunesi's Translation

ST	TT
1- "Why don't you cry again, you <u>little wretch</u> ?" (p. 104)	چرا دیگه گریه نمی کنی <u>بیچاره</u> ی بدبخت؟ (یونسی، ص. ۱۲۰)
2- To be dangling them all against the edges of <u>the stairs</u> (p. 68).	آنها را هر دم به <u>زمین</u> می کشیدم (یونسی، ص. ۶۷).
3- "Now?" she said, "you little <u>coarse monster</u> what do you think of me now?" (p. 104).	پرسید: «حالا چطور؟ بی تربیت <u>ایکبیری</u> . حالا چه جور آدمی <u>هستم</u> ؟» (یونسی، ص. ۱۲۰)

Table 12

Breaches of TL System in Yunesi's Translation

ST	TT
1- Joe kept a journey man at weekly wages whose name was Orlick (p. 131).	جو کارگر <u>هفته مزدی</u> را به نام اورلیک به مزدوری گرفته بود. (یونسی ص. ۱۵۷)
2- The coffee-room at the <u>Blue Boar</u> was empty (p. 228).	اتاق قهوه خوری <u>مهمان خانه</u> خالی بود. (یونسی ص. ۲۳۲)

When breaches of the TL system occur, it means that the translation contains errors, inconsistencies, or awkwardness in terms of target language usage. In Yunesi's translation, the number of breaches of TL system was partly the same as that in Akhavan's translation but it did not affect the content and faithfulness of the text.

4.5. Implications of Cultural Filter

The cultural filter plays a significant role in House's model of Translation Quality Assessment. It refers to the influence of cultural factors on the translation process and the potential impact on the quality of the translated text. House recognizes that translations are not purely linguistic transfers but also involve cultural and contextual considerations. The cultural filter refers to the way in which translators interpret and adapt the source text based on their own cultural background, beliefs, and values. This can result in a filtering effect, where certain cultural aspects may be emphasized or modified, leading to potential changes in meaning or perception.

Translators must make conscious decisions on how to adapt cultural references, idiomatic expressions, or cultural-specific concepts to ensure they are understood and acceptable in the target culture. The cultural filter also takes into account the characteristics and expectations of the target audience. Some cultural concepts, idioms, or expressions may be challenging to translate directly due to the differences between languages and cultures. The cultural filter also highlights the importance of the translator as an intercultural mediator. They need to bridge the gap between the source and target cultures, ensuring that the translated text is not only linguistically accurate but also culturally appropriate and relevant.

To remove or change a cultural element is cultural filtering which was seen in the target texts and occurred in Akhavan's translation frequently. The translator applied cultural filters to make the novel appropriate for Iranian and Persian culture and for teenager class of readers. In addition, Yunesi's translated text is not filtered culturally. The implications of the cultural filter are essential for a comprehensive assessment of translation quality according to House's TQA model. It recognizes that translations go beyond linguistic transfer and require a deep understanding of the cultural and sociocultural aspects associated with both the source and target languages.

Table 13

Comparison between ST and TT

ST	TT
1- <u>Mr. Pumblechook and Mr. Hubble declined on the plea of a pipe and ladies' society but Mr. Wopsle said...</u> (p. 59).	آقای وپسل گفت اگر آقای جو بره (اخوان، ص. ۲۹)
2- She came back with some bread and meat and a <u>little mug of bear</u> (p. 86).	او بعد از دقایقی با مقداری غذا و <u>یک لیوان آب بازگشت</u> (اخوان، ص. ۵۴).
3- And get behind one of the gates in the <u>brewery-lane</u> (p. 86).	رفتم پشت یکی از درهای <u>کارخانه ی نوشابه سازی ...</u> (اخوان، ص. ۵۴).
4- Well miss? I answered <u>almost falling over her</u> and checking myself (p. 104).	خم شده بودم و خودمو نگاه می کردم. گفتم: « <u>خب چی خانم؟</u> » (اخوان، ص. ۶۰).

4.6. Summary of findings

The provided data shows that 63.6% of translation errors in the study of discrepancies in the translation of Dickens's *Great Expectations* by Akhavan consist of certain parts of the text which were not translated. This can lead to some important information being left out and a loss of coherence in the translation. Mostly, it is done because of localization, cultural filters, and making a short story. The solution for this type of error is for the translator to carefully review the text and make sure all parts are accurately translated.

Addition errors account for 4.8% of the discrepancies, where the translator adds information that was not present in the original text. This can lead to a distortion of the original meaning and should be avoided. The solution for this error is for the translator to stick closely to the original text and only add information for clarity when necessary. In Akhavan's translation, addition is to make the story clear for a special class of children's ages.

Errors related to distortion of meaning make up 15% of the discrepancies, where the translation changes the meaning of the original text. This can lead to a misinterpretation of the author's intentions and should be corrected. The solution for this error is for the translator to carefully consider the context and style of the original text to ensure that the translation reflects the intended meaning accurately.

Table 14

The Percentages of Overt Errors in Akhavan's Translation

Over Errors Classifications	The Percentages of Overt Errors in Akhavan's Translation
Not translated	63.6%
Additions	4.8%
Distortion of Meaning	15%
Breaches of TL	0.8%
Slight changes in meaning	15.8%
sum	100%

Breaches of target language (TL) norms account for 0.8% of the errors, where the translation does not adhere to the conventions of the target language. This can lead to confusion for the reader and should be avoided. The solution for this error is for the translators to familiarize themselves with the norms and conventions of the target language and adhere to them in their translation.

Lastly, slight changes in meaning make up 15.8% of the discrepancies, where the translation introduces minor changes in the original meaning. While these errors may be less severe, they can still impact the overall interpretation of the text. The solution for this error is for the translator to pay attention to details and strive for an accurate and faithful translation of the original text.

Elements which were not translated (25%): This means that a quarter of the translation does not have an equivalent in the target language. This is due to the fact that the

translator missed some parts of the original text to localize them. But in some cases, it could be due to not being able to find appropriate equivalents in the target language. To improve this, the translator should carefully review the entire text and make sure to translate all parts accurately.

Table 15

The Percentages of Overt Errors in Yunesi's Translation

Overt Errors Classifications	The Percentages of Overt Errors in Yunesi's Translation
Not translated	25%
Additions	25%
Distortion of Meaning	32.7%
Breaches of TL	5.8%
Slight changes in meaning	11.5%
sum	100%

Addition (25%): This refers to instances where the translator has added information or words that were not present in the original text. This could lead to a distortion of the author's intended meaning. To address this, the translator should stick closely to the original text and avoid adding any unnecessary information.

Distortion of meaning (32.7%): This indicates that over a third of the translation does not convey the intended meaning of the original text accurately. This occurred due to some semantic errors, misinterpretation of the source text, or lack of linguistic proficiency. To rectify this, the translator should carefully analyze the original text and make sure to convey the author's intended meaning accurately in the target language.

Breaches of TL (5.8%): This refers to instances where the translation violates the rules or conventions of the target language, which can lead to confusion or misunderstanding for the reader. To improve this, the translator should have a good understanding of the target language and adhere to its grammar, syntax, and stylistic conventions.

Slight changes in meaning (11.5%): This means that there are minor alterations in the meaning of the translation compared to the original text. While this may not impact the

overall understanding of the text significantly, it is important for the translator to strive for accuracy in conveying the author's intended meaning. To address this, the translator should carefully consider the nuances of the original text and make sure to capture them accurately in the target language.

Table 16

The Percentages of Overt Errors in Yunesi's and Akhavan's Translations

Translations	The percentages of overt errors
Overt errors in Yunesi's translation	35.3%
Overt errors in Akhavan's translation	64.7%

The comparison between Yunesi's and Akhavan's translation reveals some similarities and differences in terms of the translation quality.

Similarities:

- Both translations have instances of some parts of the text which were not translated, though Yunesi has a lower percentage of these instances (25% compared to 63.6% in Akhavan's translation).
- Slight changes in meaning are present in both translations, with Yunesi having a slightly lower percentage (11.5% compared to 15.8% in Akhavan's translation).

Differences:

- Yunesi has a higher percentage of addition of the text (25% compared to 4.8% in Akhavan's translation), indicating a tendency to add information during translation.
- Yunesi has a higher percentage of distortion of meaning (15% compared to 32.7% in Akhavan's translation), suggesting that the translations in this category may deviate more from the original text.
- Yunesi has a higher percentage of breaches of TL (5.8% compared to 0.8% in Akhavan's translation), indicating more instances where the translation does not reflect the target language norms accurately.

Overall, these results suggest that Yunesi may have a slightly better translation quality because of the lower percentage of elements which were not translated, but also have a tendency to add information and make breaches of target language norms. Akhavan, on the other hand, has a higher percentage of elements which were not translated, and slight changes in meaning. Translators in both translations must prioritize preserving the integrity and accuracy of the original text while ensuring that the translation is fluent and natural in the target language.

4.7. The Chi-Square Results

To evaluate the results of the study based on House's Model of TQA, a Chi-square test can be employed. The Chi-square test is a statistical test that helps determine if there is a significant association between two categorical variables. In this context, the Chi-square test can be used to analyze the relationship between the translation quality assessment provided by House's Model of TQA (with categories such as linguistic, functional, and cultural aspects) and the actual observed translation quality in the research data. Table 17 shows the result of this test: here is an association between the translation quality assessment based on House's Model of TQA and the observed translation quality.

Table 17

The Chi-Square Results

Test Statistics	
	Overt errors in Yunesi's and Akhavan's translations
Chi-Square	31.425 ^a
df	1
Asymp. Sig.	.000

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 89.5.

Based on the Chi-Square calculation results, the P-Value is less than 0.05. Thus, it can be concluded that there is an association between the translation quality assessment based on House's Model of TQA and the observed translation quality.

5. Discussion and Conclusion

The purpose of this article was to apply House's Model of Translation Quality Assessment (TQA) to assess the quality of two Persian translations of Charles Dickens's *Great Expectations*, completed by Yunesi and Akhavan. House's TQA model provides a comprehensive framework to evaluate translation quality, taking linguistic, functional, and cultural aspects into account.

The research findings are presented and interpreted based on the analysis. The researcher draws conclusions and discusses the implications of findings in relation to the research question and objectives. Dastum's *Great Expectations* translated text was another article that is about the comparison of source text and target text, based on House's model of TQA (Mozaheb & Salmaki, 2020-2021). According to the results, it was a covert translation because it was culturally translated with great attention to the target language. The existence of overt errors and cultural filters expressed that the translated text was covert (Charkhtab & Monsefi, 2019; Heidari Tabrizi et al., 2013). The researcher showed that the cultural filter in Akhavan's translation is a strong reason that the text was covertly translated based on House's model of TQA. In Mossop's cultural filter, implication was the cultural adaption. While, the most used implication of the cultural filter in this study was removing different cultural elements and adapting (Mossop, 2003). The study by Kamala proved that the filtered cultural elements occurred because of ensuring the comprehension of the target reader (Kamala, 2007). In another study by Wu, the complexities and difficulties for translating a literary text by applying cultural filters were defined (Wu, 2012). The researcher claimed that cultural filtering was appealed to make target text appropriate for teenage class of readers. In Chen's (2015) paper, it was asserted that analyzing both the source text and target text, and then comparing them, could be a crucial method for evaluating the translation quality. This approach helps to identify the strengths and weaknesses of the translated texts, providing strong evidence for the effectiveness of House's TQA model. Based on the application of House's TQA model, both translations by Yunesi and Akhavan demonstrated a satisfactory level of translation quality. Yunesi's translation stood out for its formal accuracy and cultural

faithfulness. Akhavan's translation, on the other hand, displayed a more contemporary and reader-friendly approach, enhancing the reader accessibility.

In this study, the researcher tried to distinguish two Persian translations of Dickens's *Great Expectations* which were translated by Yunesi and Akhavan based on House's model of TQA (1997). The functional analysis assessed the effectiveness of the translations in conveying the intended purpose and tone of the source text. Both translations demonstrated a commendable level of accuracy in maintaining the plot and structure of the original work. However, Yunesi's translation was observed to be more faithful to the formal and literary aspects of the source text, while Akhavan's translation exhibited a slightly more modernized and colloquial approach. With due attention to the results and percentages, Yunesi's translation was more appropriate than Akhavan's translation based on House's model of TQA in order to convey the core meaning of the original text. Akhavan's translation was full of culturally filtered elements to be proper for teenage class of readers but in Yunesi's translation there were not any culturally filtered elements. Based on House's model of TQA (1997), Akhavan's translation was a covert translation and Yunesi's translation was an overt translation. The application of House's Model of TQA on the two Persian translations of Charles Dickens's *Great Expectations* revealed that both Yunesi and Akhavan produced translations that meet acceptable standards of quality. The translations were evaluated from linguistic, functional, and cultural perspectives, and while some discrepancies were identified, overall, the translations succeeded in conveying the essence of the source text to Persian readers. This study highlights the significance of considering House's TQA model for comprehensive translation quality assessment and emphasizes the need for ongoing evaluation and feedback to further enhance the translation excellence. Here are some suggestions for further studies based on this article:

- 1) Conducting a comparative analysis of translations of *Great Expectations* by other translators using House's Model of TQA. This will help provide a broader perspective on the effectiveness and applicability of the model in evaluating different translations.

- 2) Supplementing the quantitative analysis with qualitative methods such as interviews or focus groups to explore the perceptions and opinions of readers or target language speakers regarding the quality of the translations. This can provide valuable insights into the subjective aspects of the translation quality.

Conflict of interest

The author(s) certify/certifies that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in the present research paper.

References

- Akhavan. (1996). *Dickens's Great Expectations*. Nahal-e-Navidan. Doi: <https://doi.org/10.1007/978-1-349-07478-5>.
- Charkhtab, M., & Monsefi, R. (2019). *Application of House's Translation Quality Assessment on Children's Illustrated Book: the Case Study of Tyrone the Big Bad Bully*. The 7th National Congress on Applied Research Language Studies. Tehran. Doi: https://doi.org/ELSCONF07_095.
- Chen, Y. (2014). An Empirical Study of Cultural Filters in Chinese-English Translations: A TQA Approach. Doi: <https://doi.org/10.1080/0907676X.2014.948888>.
- Dickens, C. (2005). *Great Expectations (Barnes & Noble Classics Series)*. Orion Press Inc.
- Halliday, M. A. (1987). Spoken and written modes of meaning. In *Comprehending oral and written language* (pp. 55-82). Brill.
- Heidari Tabrizi, H., Chalak, A., & Taherioun, A. H. (2013). Assessing the Quality of Persian Translation of Orwell's Nineteen Eighty-Four Based on House's Model: Overt-Covert Translation Distinction. *International Journal of Foreign Language Teaching & Research*, 1 (2), 11-21. Doi: <https://doi.org/10.4312/ala.4.3.42>.
- House, J. (1977). *A model for translation quality assessment*. TBL-Verlag Narr.
- House, J. (2014). *Translation quality assessment: Past and present*. Routledge.
- Kamala, S. (2007). The translation quality of cultural references in literary texts: an application of House's model of TQA. Doi: <https://doi.org/10.36892/ijlts.v4i3.349>.
- Masoudnia, N. (2017). The Application of House's TQA Model on Paula Hawkins's "*The Girl on the Train*" and its Two Persian Translations by Faranak Salary and Mahbobeh Mosavi. Shahr-e-Qods, Iran: Islamic Azad University.
- Mossop, B. (2003). Towards describing translator competence. Doi: <https://doi.org/10.7202/029688ar>.

- Mozaheb, M. A., & Salmaki, M. (2020-2021). *Analysis of Dickens's Great Expectations Using House's Model of Translation Quality Assessment*. The 7th International Conference on Modern Finding in Science and Technology with a Focused on Science in Development Service. Tehran. Doi: https://doi.org/DSCONF07_121.
- Wu, J. (2012). *Assessing the Quality of Covert Translations: A Comparative Study based on House's TQA Model*. Lambert Publishing.
- Yunesi. (2020). *Dickens's Great Expectations*. Doostan. Doi: <https://doi.org/10.1007/978-1-349-07478-5>.

Similarities and Differences between News Covering the Same Event in Persian and English News Agencies: the Skopos Theory Perspective

Pantea Changizi 

PhD candidate in Translation Studies, Translation Department, Faculty of Literature and Foreign Languages,
Allameh Tabataba'i University, Tehran, Iran.
panteachangizi@yahoo.com

Abstract

This paper aimed at investigating the similarities and differences between the news of the same event published in Persian and English languages. Persian was the source and English was the target language. By target language it may come to the mind that the English news agency websites are the translational versions of the same Persian news agencies, but this is not true in the case of the current paper. English news websites were considered as tools and channels for spreading the news regarding the Iranian Metropol building collapse, so foreigners would notice to events that happened in Iran. The aim of using the Skopos theory viewpoint in this study was to investigate the reason for the existing dissimilarities found between English and Persian news. Based on the Skopos theory, it was concluded that the dissimilar parts of each language in reporting and covering the Metropol building collapse news, resided in the guidelines which were dictated to news writers to write and publish on their news agency websites. The points covered in this paper can be an additional point of departure from a mere theoretical study of the Skopos theory.

Keywords: Skopos theory, News, Translation, Persian, English

Citation: Changizi, P. (2024). *Applications of Language Studies*, 2(2), 254-274.

1. Introduction

When we talk about translation and media, the first thing which may come to our mind as a kind of media is news. However, nowadays the way in which news are grasped and understood is much different from the traditional viewpoint. In contrast to the past, today people do not solely get the news from TV, rather they may hear or read about it on the Internet or any other new different way. In other words, news is no longer limited to the television screen and every day, every second we are all surrounded by dozens of news published everywhere in the world.

Whether the news is published through television, newspaper, or social media, it does not make a difference in the fact that they are all labeled as one thing called 'news'. Also considering the word translation, when it comes to its merge with news and generally the news translation, varied factors can be turned into a possible area of study. In this paper, we aimed at exploring the differences between published news about a same event in English, from the Skopos theory viewpoint.

2. Review of Literature

2.1. Skopos Theory

Translation of a text can be studied beyond its textual and grammatical levels. Instead of just focusing on points like translational equivalence or problems, you can go further than it and consider and study things like contexts surrounding that text. By context it does not mean the textual context, rather, this goes beyond the text and regards things like cultural, social, political, and other contextual dimensions.

Here is where the Skopos theory comes. Skopos theory is an approach to translation which was developed in Germany in the late 1970s, and which reflects a general shift from predominantly linguistic to a more functionally and socio-culturally oriented concept of translation"Invalid source specified.. Passing the equivalence stage of a text between two different languages, according to the Skopos theory, the focus shifts toward the concept of 'purpose' behind each text. Function is yet another term that refers to what a text means.

Another related term to Skopos is intention which is regarded as an aim- oriented plan of action on the part of both the sender and the receiver” (Kadhim Jabir, 2006, p. 38). Each text in any language serves a purpose and function. Depending on societies’ norms and conventions, there are some guidelines for the production and distribution of texts whether published as a hard copy or in an electronic format.

The top-ranking rule is the “Skopos rule”, “the end justifies the means”. Vermeer explained the Skopos rule as follows: Translate/interpret/speak/write in a way that enables your text/translation to function in the situation in which it is used and with the people who want to use it and precisely in the way they want it to function. Invalid source specified.

According to the aforementioned points, it can be said that through the lens of Skopos theory, it is the target society which determines how and in which way and with what features a source text should be translated or generally transmitted. In a simpler view, consider news published in Iran in Persian and now it is going to be translated in English (a target culture). According to Skopos theory, target culture and society in which the language is English, dictate how Persian news should be translated. That is to say, which parts of the source text (Persian news) must be eliminated or even be bolded.

The reverse is true if English news is going to be translated into Persian, it is the Persian society which decides for the translational features to be considered while doing the translation. In this paper, in translation of news from one language (source) to another (target), the focus is not on translational action in its linguistic definition (e.g., which words are exactly or freely translated from language A to language B).

The aim of this paper is to emphasize on transport, transmission, and conveyance of news from a source language (Persian) to a target language (English). Thus, it could be better to use the word transfer rather than translate since the purpose is to study the similarities and differences of news of a same event occurred in Iran in Persian and English news agency websites which have covered that specific news.

2.2. Skopos theory research in Iran

Some researchers have worked on the theoretical aspect of the Skopos theory like a study (as an example) in which the researcher has worked on this theory and studied some of its aspects along with studying Jean-René Ladmiraal theory (Zendebudi, 2012).

As the second area in the Skopos theory research in Iran, a work is done by Haji Abdollahi and Athaari Nik Azm. They have used the Skopos theory to study the translation of a movie subtitles. They have analyzed the Persian subtitles of a French movie called ‘Class’. Among important criteria for the quality of a subtitle, they assumed that the coordination of picture and voice, conciseness, clarity and simplicity of the sentence structures, are the most important features (Athaari Nik Azm & Haji Abdollahi, 2015).

The next study belongs to Afzali and Shahbazi, who have worked on the use of Skopos theory on assessing a translation of Quran by Ali Maleki. This study has been carried as a case study on the Al Modasser sura. In this study, based on an analytical-descriptive method, they have studied the translation of the aforementioned sura from the book: ‘The Holy Quran, readable translation of the Holy Quran in an interpretive way for adults and young adults’. Then, they have applied the Skopos theory in order to show the degree of the translator’s success in achieving the purpose of the translation based on the Skopos theory. The findings showed that the translator has not been loyal to the ST in some considerable aspects; however, he has produced a very coherent and comprehensive translation (Afzali & Shahbazi, 2022).

Mahdavi Ara and Noori Keyzghani have used the Skopos theory on studying Ghesas Al Hayvan translation for children and young adults. They have studied the functionality of the Skopos theory in children and young adults’ literature translation, and have studied Ghesas Al Hayvan Persian translation with a critical approach based on the Skopos theory (Mahdavi Ara & Noori Keyzghani, 2022).

The above-mentioned names and research studies are included as a very short list of how and in which areas the Skopos theory has been studied and worked on by Iranian researchers. It is clear that this short list is not all-inclusive at all and there might be more

studies done with an interest in Skopos theory. But the thing which is apparent in these examples is that they mainly have focused on the translational aspects of one work (whether a movie or a book) and considered the role, aspects, and representation of the Skopos theory in them.

In the above-mentioned short list, no study has compared two different languages (news in two different languages; one language as a source and the other as the target) in order to find the similarities and differences existing in them based on the Skopos theory.

3. Method

The current paper aimed at studying news (from an Iranian incident) published on late spring 1401 in Iran; the collapse of Metropol building in the south of Iran, Abadan. For this paper, some news of this event was randomly chosen from news agency websites (see the appendix). Five random English language and five random Persian news agency websites were chosen. The common feature between the Persian and English news agencies was only the period of time during which the news was published. It was tried to consider the time period from the late spring 2022 (based on Persian calendar; from Khordad 1401 until Aban 1401) until the beginning of the autumn 2022.

From the five random English news agencies and five random Persian news agency, the body of the news was studied in order to find the similarities and differences between the Persian news which is considered as the source text and the English news which is the target text (target in a way that the English news agencies were considered responsible as a communicational channel and tool for conveyance of that specific Iranian incident to the foreigners and other nations and countries so that they would be noticed about the Metropol building collapse through the English version of the news). These two languages were compared and the unit of comparison was a sentence. The sentence unit was considered as a separate sentence wherever there existed a dot.

For a better understanding of the aim of the current study; which focuses on showing the similarities and differences between two languages of Persian and English used in news agency websites for the coverage of the Metropol building collapse occurred in Iran in late

spring 2022, some sentences of each of five English and five Persian news agency websites are extracted and put in separate tables which will be presented in the following sections. Similar parts of the news in both languages are bolded. These similar parts are not exactly similar (like a translational equivalence), rather, they both convey approximately the same aspects of the Metropol building collapse.

The dissimilar parts are in italics. These dissimilarities are mainly chosen in order to highlight the differences of two languages in the way of reporting and representing the Metropol news based on the Skopos theory that determines how the news should be published on websites due to norms and regulations of each society. In the following section, first, five English news tables are given with their extracted sentences. Then, five Persian news tables with their extracted sentences are provided.

4. Results and discussion

In this section the results of the analysis will be presented and the similarities and differences between the source and target versions of the news will be revealed.

4.1. Similarities

As the boldened parts in both languages show, there are some similarities between the news agency websites in their news coverage. Both of the languages have referred to the following points:

1. Number of the dead and injured people;
2. Reason of the collapse due to the neglect of supervisors and related authorities to the warnings;
3. Low quality of the materials used in the building;
4. Arrested suspected people;
5. Ongoing work of rescue teams.

4.2. Dissimilarities

The dissimilarities between the English and Persian news agency websites which are in italics can be summarized into the following points. The English websites have referred to the points which were not seen in the five chosen Persian news websites:

1. Unknown number of people being trapped;
2. Continuous crisis and disaster of building collapse in Iran;
3. Metropol reminding of Plasco building collapse and other recent similar building catastrophes in Iran;
4. Metropol news as a “top breaking-news topic” beyond the Iranian borders;
5. Blaming of Iranian building construction standards;
6. Repetition of building collapses in Iran;
7. Lack of care of authorities toward such incidents;
8. Naming some reasons for Metropol collapse;
 - a. Increase of population and high cost of houses
 - b. Inflation in Iran
 - c. High price of building construction materials
 - d. Preference of low quality but cheap materials
 - e. Low quality materials resulting in death of people and money making for constructors
9. Putting in danger the life of citizens;
10. Criticisms and protests and shouting of people due to such building constructions;
11. Unequipped and weak rescue teams;
12. mentioning death of two policeman;
13. 80 people being still trapped;
14. Numerous equipment being sent from other cities to help rescue teams in Abadan (naming them with detail);
15. Some people watching the process of rescue team “in horror and desperation”;

16. English news agency (number 5) giving warning and safety advises.

Now consider the dissimilarities of the five Persian news websites in comparison to the five English websites:

1. Applause of rescue teams and asking for their appreciation;
2. Asking people to remain calm and not to attend at the incident site;
3. Naming some students killed at this incident;
4. Mentioning specifically the students' names;
5. Condolences for the death of mentioned students;
6. Naming the owner of Metropol building and referring to his unclear news of death;
7. Naming Khatam Al Anbia as being designated for destruction of the remained building;
8. Naming Khatam Al Anbia as a self-considered responsible organization for safety of people;
9. Five months after the collapse and a re-collapse of the remained building;
10. Mourning and gathering of people in memory of the dead ones;
11. Mentioning the sentence and verdict of the court for the accused people;
12. Fast and high-speed process of investigations;
13. Putting emphasis on the speedy process of investigations;
14. Personal visit of the attorney general from the victims and asking for their needs and promising to meet their needs.

One point should be mentioned here that among the five Persian agency news websites, only one of them (Persian news number 4) has only mentioned this incident as a catastrophe (فاجعه) and has blamed and criticized heavily (in comparison to other Persian

websites) the related authorities or organizations and it also has referred to the protest of people. These parts have been underlined in the table related to Persian news number 4.

Going back to comparing the similarities and differences between these two languages in news websites, first we discuss the similar points. As the bolded parts in tables and numbered bullet points show, these similarities in both Persian and English news websites are quite near to each other in terms of representing the general aspects and coverage of the Metropol building collapse.

But the dissimilarities in both languages are not that much general and they refer to more specific aspects of the Metropol news. For instance, in five English news, the aim is to shed more light on the lack of responsibility and weakness of the Iranian authorities, weak equipment of the rescue team, and the rage and wrath of people and their protests. On the contrary, the five Persian news websites (except an exceptional part in Persian news number 4 which was mentioned above) have covered points which are more relevant to a domestic receiver of the Metropol collapse news rather than a foreigner one. For instance, an Iranian news reader would be more interested in following the final verdict of the court in comparison with a foreign person who lives miles away from this incident. The Persian news has appreciated the efforts and works of the rescue teams and has emphasized on the fast speed of the investigation process.

From these similarities and dissimilarities, it can be understood that the dissimilar parts can clearly show the policy and guidelines of each country's news agency websites. The policies of an English news agency dictate to write the news in a way to represent a positive or negative picture from another country based on their political stance and relations with that foreign country (whether Iran or any other countries).

On the other side, the Iranian news agency has to report the situation in a way to both cover the news and also to try to control the turbulent atmosphere of the country. So if the domestic news agency emphasized on issues including lack of proper speed or weak equipment, or try to repeatedly criticize the related authorities and organizations, it would provoke and inflame people's wrath which might result in more chaos and disorder especially near the collapsed building zone.

Considering the guidelines and preferences of news agency websites of each language, these are what have been the focus and core point in Skopos theory. As mentioned in previous parts of this paper, it is the Skopos of each society and purpose of it which dictate the translator, news writer, content developer, or any other person who is responsible in producing information for the audience. In this paper, the English news websites are the communicational channel (in some way translational but in terms of transmitting the news) through which the target people (foreign people of other countries) become informed about what has happened in another country (Iran).

So according to the Skopos theory, “the prime principle determining any translation [or information transmitting] process is the purpose (Skopos) of the overall translational [communicational] action” Invalid source specified. that is the target society which tell how the source text has to be transferred into the target language. The culture, the politics, and many other factors of the target society are important in using Skopos guidelines.

Generally speaking, whenever there is news published somewhere on this planet, when it comes to its transference among different countries, each country covers that news in accordance with its political relations with the source language society and interpret that news for its people in the way that the purpose or Skopos of that country dictates to it. So it can be understood that this is a trend and reoccurring tendency in the sphere of news publishing; each country should obey norms, rules, and conventions which are given to it from its own society.

5. Conclusion

The Skopos theory is one of those theories which can be used as a tool in studying target societies guidelines which are applied and obeyed in the process of translating or transferring the information or news from a source society to a target society. For instance, if we are going to study a text and go beyond the grammatical and lexical levels and go through its cultural context and focus on the purpose behind publishing that specific text in the target society, we can use the Skopos theory as a useful tool. In this sense, the Skopos theory tells us about what the target society and culture impose based on its political, cultural, sociological, and other existing norms and conventions. As long as these features does not

distort the originality of the news and what has really happened, there will be no problem. However, if they cause a radical change in the nature of the news and distort it in a way resulting in a totally radical different interpretation of that news on the part of the news readers, there will be undesirable consequences.

Conflict of interest

The author(s) certify/certifies that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in the present research paper.

References

- Afzali, A., & Shahbazi, L. (2022). Use of the Skopos theory in assessing the quality of Ali Maleki's translation of the Holy Quran; A case study of Al Modasser Sura. *Islamic Literature Studies*, 1(2), 163-185. DOI: 10.34785/J022.2022.016.
- Athaari Nik Azm, M., & Haji Abdollahi , M. (2015). Audiovisual translation, studying the Class movie subtitle based on Functional Theories of Grammar and the Skopos theory.
- Dan, J. (2015). An overview of German's functionalist Skopos theory of translation cognition. *Studies in English Language Teaching*, 3(3), 203-214.
- Du, X. (2012). A Brief Introduction of Skopos Theory. *THEORY AND PRACTICE IN LANGUAGE STUDIES*, 2(10), 2189-2193. DOI: 10.4304/tpls.2.10.2189-2193.
- Hashemi, M., & Heydarpoor, A. (2013). Studying the Skopos theory (functionality) on the Baghdad Translation Movement.
- Kadhim Jabir, J. (2006). Skopos theory: Basic principles and deficiencies. *Journal of the College of Arts*, 41(2) 37-44.
- Mahdavi Ara, M., & Noori Keyzghani, S. (2022). Studying the functionalities of the Skopos theory in children and young adult literature translation (a case study of Persian translation of Ghesas Al Hayvan from the Holy Quran).
- Schäffner, C. (1997). Skopos theory. In M. Baker, & G. Saldanha (Eds.), *Routledge encyclopedia of translation studies* (pp. 235-238).
- Suo, X. (2015). A new perspective on literary translation strategies based on Skopos theory. *Theory and Practice in Language Studies*, 5 (1), 176-183. DOI: <http://dx.doi.org/10.17507/tpls.0501.24>.

Zendehboudi, M. (2012). Skopos Theory in Translation Studies. *Iranian Journal of Translation Studies*, 9(36). Retrieved from <https://journal.translationstudies.ir/ts/article/view/1078>.

Appendix: Persian and English news

Table 1

English News 1

English news 1 ¹
Building collapse death toll in southwestern Iran rises to 26.
An additional seven bodies have been pulled from the rubble of a collapsed building in southwestern Iran, bringing the death toll from the disaster to 26.
Rescue teams at the site of collapsed building in southwestern Iran pulled seven more bodies from the rubble on Friday, bringing the death toll in the disaster to 26 , Iranian state TV reported Friday.
Thirty-seven people were rescued, three of them are still being treated in hospital. <i>It remains unclear how many more people are buried beneath the rubble.</i>
Authorities have arrested 11 suspects in a widening probe, including the city's mayor.
The deadly collapse has raised questions about the safety of similar buildings in the country and <i>underscored an ongoing crisis in Iranian construction projects that has seen other disasters in this earthquake-prone nation.</i>
Abadan provincial officials have cited “disregard for technical standards” and “overbuilding” during construction, saying negligence caused the collapse.
The building was legally permitted to be only a six-story tower, but four floors had been added during construction.
The collapse <i>reminded many of the 2017 fire and collapse of the iconic Plasco building</i> in the Iranian capital, Tehran, which killed 26 people.

Table 2

English News 2

English news 2 ²
<i>The memory of the collapse of the Plasco building in Tehran in 2017 has not been erased from the Iranian minds (this event was one of the most shocking technological disasters in Iran) when, another catastrophic incident, the collapse of the newly constructed Metropol building in Abadan city in 2022, occurred.</i>
<i>The dimensions of this incident were so broad that it was at the top of the breaking news in national and international media.</i>
As of June 7, 2022, a total of 43 people were killed and 37 injured in this incident.
This collapse was due to structural failure due to a violation of technical regulations and standards.
Thirteen of the culprits of this incident have been arrested.
The instability of the structure was the main cause of this incident.
The collapse of the Metropol building marks <i>the continuous occurrence of similar catastrophic incidents within building construction</i> in the country, despite the special emphasis of the authorities and institutions on following up and complying with the standards of building construction.

¹ <https://www.newarab.com/news/building-collapse-death-toll-southwest-iran-rises-26>

² <https://www.cambridge.org/core/journals/disaster-medicine-and-public-health-preparedness/article/collapse-of-the-metropol-building-recurring-manmade-disasters-in-iran-2022/EFE3581290F062DAE04BA8B466E2BC85>

This raises the following questions: *What standards do building construction and contractors, in general, follow in Iran?*

Why does the collapse of buildings in Iran keep repeating despite the emphasis on standardization and prosecution of the culprits of these incidents?

What institutions, persons, and officials should be accountable to the people in this field?

According to experts and city officials, *the possible causes of the Metropol building incident are:*

The increase and density of the population in big cities, the high cost of housing and the increase in the inflation rate, the increase in the price of construction materials, and other related issues have fueled other challenges in construction and urban development in recent years.

The high cost of construction materials and the outflow of private sector capital have unpleasant consequences in the construction of residential and non-residential structures.

*As a result of the high cost of building materials, contractors, and construction companies seek to reduce the costs of building structures, and often regardless of the consequences of this decision, they **consider the lowest budget for building construction** and get the most benefit; The result of this action is the use of **low-quality materials, nonstandard structures**, and *catastrophic incidents, which ultimately cost not only people's pockets but also people's lives.**

Neglecting technical and engineering standards in construction, including the use of low-quality, nonresistant, and nonstandard materials, lack of supervision during the construction of buildings, and not following legal procedures in issuing building permits are among the things that lead to the creation of **nonstandard structures**.

Abadan Metropole building is one example, **where the legal procedure for issuing an operating permit has not been followed**.

According to the head of the Khuzestan Province Building Engineering System Organization, **the municipality was informed about "flaws such as non-compliance with the rebar cover, the deformation of some beams, and the buckling of some columns of the Metropole building project 1 month before the incident, and the order to stop the project was given, but the order was not implemented"**.

Despite the numerous regulations and directives regarding building construction; **the construction of unsafe and nonstandard buildings, the use of cheap and low-quality materials**, nonresistant and disproportionate to the weight of the structure, along with the **negligence of the relevant authorities** to control all technical circumstances before issuing construction and operating permits, *has resulted in serious challenges in building safety, jeopardizing the safety of citizens.*

*Numerous incidents, especially in the past 5 y, such as the Plasco building incident in 2017, the Sina Clinic fire in 2019, and the collapse of the Abadan Metropol building in 2022; apart from the **lack of compliance with technical and engineering standards**, highlights the **lacking supervision and accountability of the relevant authorities and managers.***

The disaster of the collapse of the Metropol building has *once again increased the criticism of the construction process in the country and the possibility of subsequent similar incidents.*

This painful disaster shows the ineffectiveness of the monitoring system in safe construction.

It is still too early to make a final judgement, *but citizens have the right to ask: Why has no action been taken to stop the construction despite the warnings (A few months before the incident)?*

Why did the rescue forces arrived late and not have enough equipment and dozens of other questions that prevent the "Metropole Tower" incident in Abadan erasing from the memory of Iranians anytime soon.

Table 3

English News 3

English news 3¹

Tehran (AFP) – Five people including two policemen were killed and at least 11 others injured during the collapse of several buildings in Iran's capital, local media reported on Monday.

Rescue operations were underway to find *others who may be trapped* under the rubble, ISNA and Tasnim reported.

Photographs from the scene issued by Tasnim showed **rescuers and heavy equipment working** below a pile of debris and twisted metal several metres (yards) high, with other unfinished and partially damaged buildings around them.

ISNA quoted a police statement saying the **buildings which collapsed did not comply "with construction safety measures"**.

A Tehran city official said on Saturday **that authorities had demolished more than 46,000 unauthorised buildings over the past two years**, according to Hamshahri newspaper.

The collapse of the 10-storey Metropol building, which had been under construction in the city of Abadan in Khuzestan province, *sparked protests across the country decrying corruption and incompetent authorities.*

Table 4

English News 4

English news 4²

Local officials say up to *80 people are feared trapped* under the rubble after a 10-story building collapsed in the city's busiest street.

Large parts of the commercial Metropol building, located in the busiest street at the centre of the city, were reduced to rubble, and officials told state television up to *80 people may still be trapped.*

Rescue dogs, helicopters, vehicles, and personnel have been mobilised from across several provinces to assist with the ongoing rescue operation.

Those include 80 forces from Tehran, namely 50 firefighters and rescuers to assist operations on the ground and 30 emergency responders to provide aerial assistance.

One clip also showed *a group of angered citizens shouting at local municipality officials present at the scene.*

Others were watching the rescue operation in horror and desperation.

President Ebrahim Raisi, who was in Oman for a state visit at the time of the incident, in addition to his first vice president, **Mohammad Mokhber, issued orders for local officials to immediately devote their resources to helping rescue operations.**

Inspectors had reportedly **issued safety warnings months prior to Metropol's collapse.**

¹ <https://www.france24.com/en/live-news/20230807-iran-building-collapse-death-toll-rises-to-five-media>

² <https://www.aljazeera.com/news/2022/5/23/five-dead-after-building-collapses-in-irans-abadan>

Table 5

English News 5

English news 5 ¹
Sections of the building initially collapsed in May, killing 43 people.
Rescue operations are ongoing.
<i>Lingering disruptions are possible after crews clear the site as officials conduct investigations.</i>
<i>Avoid the area until the situation stabilizes and authorities give the all-clear.</i>
<i>Heed all official transport and security instructions, including any evacuation orders.</i>
<i>Plan for localized traffic congestion in nearby areas; allow additional time to reach destinations.</i>
<i>Consider remaining indoors if dust is heavy, especially if you suffer from a respiratory illness such as asthma, or conditions that lead to increased risk from dust exposure.</i>
<i>Eye irritation is possible; flush irritated eyes with copious amounts of clean water.</i>

Table 6

Persian News 1

Persian news 1 ²
عضو شورای شهر تهران با اشاره به اینکه در اسفند ۹۸ مجوز افزایش طبقات متروپل تا ۱۲ طبقه صادر شده بود گفت: با توجه به اینکه سه طبقه بدون نقشه مقاومسازی اضافه شده بود، در کمتر از یک ماه از صدور هشدار، فرو ریخت.
عضو شورای اسلامی شهر تهران ضمن تقدیر از تلاش آتش نشانی و حضور ۵۰ نیروی این سازمان در متروپل تصریح کرد: آتش نشانان به صورت سه شیفت و شبانه روزی در محل مشغول به کار بودند و اگر در آن محل حضور نداشتند مشخص نبود که خارج کردن فوت شدگان از زیر آوار به چه مرحله ای برسد؛ لذا از هیئت رئیسه انتظار داریم تقدیر ویژه ای از آتش نشانی داشته باشد.
وی ادامه داد: اصرار مالک در این ساختمان تغییرات پی در پی در طرح بوده است و بنابر اعلام نماینده شرکت نقشیداد که در ابتدا اجرای طرح را بر عهده گرفته بود، این ساختمان در آغاز با اسکلت فلزی و در شش طبقه طراحی شده بود که در ادامه بر اساس اصرار مالک به سازه های بتنی تغییر کرد که با توجه به این موضوع، شرکت نقشیداد از ادامه پروژه انصراف داد.
آقامیری با بیان اینکه ایمنی، صاحب صلاحیت نبودن طراحان طرح و استاندارد نبودن سازه ها از جمله عواملی بوده که طرح را مسدود کرده بود، یادآور شد: مهندسان ناظر دیگری برای این طرح پیش بینی شده بود که به دلایل حقوق و مزایا و مواردی از این قبیل، ناظران انصراف دادند.
وی با اعلام اینکه در اسفند ۹۸ مجوز افزایش طبقات از سوی شهرداری آبادان صادر می شود، خاطرنشان کرد: مجوز افزایش طبقات تا ۱۲ طبقه صادر شده بود و با توجه به اینکه سه طبقه بدون نقشه مقاومسازی اضافه شده بود، مشکلاتی را ایجاد می کند و اعلام می شود که ساختمان در معرض خطر اساسی قرار دارد و کمتر از یک ماه از این هشدار، متروپل فرو ریخت.
وی با بیان اینکه ضعف دانش کارفرمایان و در اولویت بودن منافع مالی برای آنان یکی از دلایل این حادثه بود، گفت: وقتی ۴۵ روز برای طراحی سازه زمان نیاز است نباید به یک یا دو هفته کاهش پیدا کند و همچنین زمان کنترل مضاعف کاهش یابد تا پروانه سریعتر صادر شود که این روند در خصوص ساختمان متروپل اتفاق افتاد.

¹ <https://crisis24.garda.com/alerts/2022/10/iran-building-collapse-in-abadan-khuzestan-province-oct-26-kills-one-person>

² <https://www.khabaronline.ir/news>

عضو شورای اسلامی شهر تهران با بیان اینکه عدم کنترل صحیح صلاحیت واقعی افراد جهت اجرا و عدم رعایت نظامات اداری در حوزه اجرا نیز از عوامل ریزش بود، گفت: عدم نظارت بر جوش تیرها و ستون ها و حرفه ای نبودن جوشکاری در این بخش ها دلیل اصلی ریزش ساختمان بود.

آقامیری با اشاره به عدم استمرار نظارت بر ساختمان در زمان ساخت و بهره برداری اظهار کرد: فسادپذیر بودن نظارت نیز از دیگر چالش هایی بود که در این پروژه با آن رو به رو بودیم.

Table 7

Persian News 2

Persian news 2¹

آبادان- فرماندار آبادان گفت: تاکنون ۱۶ نفر از شهروندان آبادانی در حادثه ساختمان متروپل جان خود را از دست دادند. به گزارش خبرنگار مهر، احسان عباسپور ظهر امروز در جمع خبرنگاران اظهار کرد: عملیات جست و جو، آوار برداری و امداد رسانی برای نجات شهروندان زیر آوار مانده احتمالی و بیرون آوردن پیکرهای جان باختگان حادثه بی وقفه و با جدیت در حال انجام است. وی در ادامه تعداد جان باختگان در این حادثه را تا این لحظه ۱۶ نفر و تعداد مصدومانی که از زیر آوار بیرون کشیده شدند را ۳۷ نفر اعلام کرد.

فرماندار آبادان با اشاره به ناپایداری شرایط ساختمان و احتمال ریزش سایر قسمت ها آن، گفت: از مردم درخواست می شود که خونسردی خود را حفظ کرده و با خودداری از تجمع در محل نیز گروه های امدادی و خدمات رسان را در انجام وظایف خود و نجات محبوس شده ها یاری کنند.

Table 8

Persian News 3

Persian news 3²

اهواز- ایرنا - سرپرست اداره کل آموزش و پرورش خوزستان گفت: در حادثه ریزش ساختمان متروپل آبادان پنج دانش آموز جان خود را از دست دادند که پیکرهای آنها از زیر آوار خارج شده است.

مسعود حمیدی نژاد روز سه شنبه در گفت و گو با خبرنگار ایرنا افزود: سه دانش آموز از خرمشهر و ۲ دانش آموز از آبادان در این حادثه جان باختند. وی با اظهار تاسف از این حادثه که به جان باختن تعدادی از هموطنان و مصدوم شدن ۳۷ نفر منجر شد گفت: "ملیکا و میترا صالحیان پور" و مسیح صادقی از خرمشهر همچنین "آرین و حمیدرضا جلیلیان" از آبادان در حادثه ریزش برج متروپل جان خود را از دست دادند.

سرپرست اداره کل آموزش و پرورش خوزستان که هفتم خردادماه در پی یافتن پیکرهای سه دانش آموز خرمشهر جان باخته در حادثه متروپل پیام تسلیتی صادر کرده بود، امروز سه شنبه نیز در پیامی درگذشت آرین و حمیدرضا جلیلیان ۲ دانش آموز آبادانی در این حادثه را تسلیت گفت.

در پیام تسلیت حمیدی نژاد آمده است: خبر تاسف بار درگذشت دانش آموزان آرین و حمیدرضا جلیلیان از آموزش و پرورش آبادان در حادثه ریزش ساختمان متروپل موجب تأثر و تألم مجموعه تعلیم و تربیت در استان خوزستان شد.

¹ <https://www.mehrnews.com/news>

² <https://www.irna.ir/news>

تعداد کشته شدگان حادثه ریزش ساختمان متروپل آبادان تا صبح روز سه شنبه دهم خردادماه تعداد ۳۴ نفر و مصدومان ۳۷ نفر می باشد.

عملیات جست و جو و آوار برداری برای یافتن پیکرهای احتمالی همچنان به صورت شبانه روزی ادامه دارد.

Table 9

Persian News 4

Persian news 4¹

چه کسی مقصر فاجعه پی در پی متروپل است؟

فاجعه متروپل ظهر روز دوشنبه ۲ خرداد ۱۴۰۱ در خیابان امیرکبیر (امیری) واقع در مرکز آبادان رخ داد.

این حادثه ۴۳ فوتی و ۳۷ نفر مصدوم داشت.

دوم خرداد ماه سال جاری بخشی از برج های دوقلوی متروپل آبادان فرو ریخت. این برج جزو یکی از هلدینگ های «حسین عبدالباقی» مالک و مقصر اصلی این فاجعه بود که پس از فرو ریختن برج متروپل مرگ او نیز با ابهام روبه رو شد.

به گزارش اعتماد، این فاجعه بر اثر شکست سازه های و نقض در ضوابط فنی و استانداردها رخ داد و در پی آن ۴۳ نفر کشته و بیش از ۳۰ نفر مصدوم شدند.

چند روز بعد در ۱۹ خرداد ماه فالح صفایی معاون فنی و زیربنایی سازمان منطقه آزاد اروند درخصوص این فاجعه اعلام کرد: «پیرو مصوبه شورای تامین استان خوزستان سازمان منطقه آزاد اروند به عنوان مدیر و کارفرما همچنین قرارگاه سازندگی خاتم الانبیا به عنوان پیمانکار تکمیل عملیات آواربرداری و تخریب بقایای ساختمان متروپل آبادان تعیین شدند.

با انعقاد این قرارداد عملیات حساس و تخصصی آواربرداری و برداشت بقایای ساختمان متروپل از سوی شورای تامین استان خوزستان به عنوان پروژه ای مستقل میان این سازمان و آن قرارگاه در دستور کار قرار گرفته است.

عملیات این سازمان فارغ از جایگاه حقوقی و چندگانگی حاکمیتی خود را در برابر جان شهروندان مسوول می داند و عملیات آواربرداری و تخریب ایمن و تدریجی ساختمان متروپل تحت نظارت سازمان منطقه آزاد اروند و از سوی قرارگاه سازندگی خاتم الانبیا طی مدت ۴ ماه انجام می شود.»

۵ ماه از فاجعه متروپل می گذرد.

نه تنها طبق مصوبه ذکر شده در این قرارداد، عملیات آواربرداری با تخریبی ایمن صورت نگرفت، بلکه برای بار سوم فرو ریخت.

چه کسی مقصر است؟

وقتی که متروپل برای بار سوم فرو ریخت گردشگران زیادی برای خرید از شهرهای مختلف خوزستان به این خیابان آمده بودند.

همین موضوع می توانست فاجعه ای دیگر به همراه داشته باشد ولی به دلیل ساعتی که این اتفاق رخ داد، تعداد مصدومان کم بود، اما متأسفانه یک خانم بر اثر ریزش متروپل جانش را از دست داد.

ساختمان تجاری پارمیدا و پارسه نزدیک ترین مجتمع ها به متروپل هستند.

این دو مجتمع در خیابان اصلی امیری نرسیده به ساختمان متروپل واقع شده اند.

برای همین مغازه دارها و کسبه های این مجتمع ها پس از این حادثه عنوان کردند که شنیده بودند متروپل قرار است بار دیگر بر اثر تخریب ویژه، ریزش کند ولی کسی به صورت رسمی این تخریب را اعلام نکرده بود.»

¹ <https://fararu.com/fa/news>

او در ادامه می‌گوید: «وقتی ۴ خرداد ماه متروپل ریزش کرد و آن فاجعه را به بار آورد، قرار شد یک کار گروهی عملیات تخریب را به عهده بگیرد. برای همین منطقه آزاد اروند پس از فاجعه متروپل تفاهمنامه‌ای را با قرارگاه خاتم منعقد کرد، اما چند روز پس از انعقاد این تفاهمنامه قرارگاه خاتم به تعهدات خود عمل نکرد.»

حادثه‌ای که ظهر روز دوشنبه ۲ خرداد ۱۴۰۱ در خیابان امیرکبیر (امیری) واقع در مرکز آبادان رخ داد.

این حادثه ۴۳ فوتی و ۳۷ نفر مصدوم داشته است.

در پی فاجعه متروپل مردم آبادان از روز دوشنبه ۲ خرداد ۱۴۰۱، در خیابان‌ها تجمع کردند و به وقوع این حادثه و کمبود تجهیزات امدادی و عدم رسیدگی به افراد زیر آوار مانده اعتراض کردند.

ساعت ۲۱ روز پنجشنبه ۵ خرداد ۱۴۰۱ تجمع عزاداری برای جان‌باختگان برج متروپل در فلکه طوطیا آبادان با حضور صدها نفر از مردم برگزار شد. همچنین ساعت ۱۹ همان روز، مردم آبادان در چهارراه شهید مطهری تجمع کردند و با شعار دادن خواستار برخورد و محاکمه قاطع و جدی با مسببان این حادثه شدند.

این ریزش بر اثر شکست سازه‌ای به خاطر نقض ضوابط فنی و استانداردها بود.

۲۱ نفر از مقصران این حادثه دستگیر شده‌اند.

در آخرین جلسه بررسی قرار شد کیفرخواست ۱۰ نفر از متهمان این پرونده مورد بررسی دادگاه قرار بگیرد.

سخنگوی قوه قضاییه اعلام کرد که هر کدام از متهمان پرونده متروپل آبادان به پرداخت دیه کامل در حق اولیای دم و پرداخت دیه مصدومان این حادثه براساس میزان تقصیر متهمان، تعیین شده است.

۲۱ نفر از جمله شهردار اسبق، شهردار زمان حادثه، سرپرست شهرداری منطقه یک آبادان و جمعی دیگر به اتهام قتل شبه عمدی ۴۳ نفر که ۷ نفر از آنان زن بودند و مصدومیت شبه عمد ۱۳ نفر تحت تعقیب قرار گرفتند؛ میزان تقصیر هر کدام مشخص شد و متعاقب آن فرآیند تعقیب و بررسی در دادسرا انجام شد و پرونده با کیفرخواست به دادگاه کیفری دو شهرستان ارسال شد و رای آن صادر شده است.

هر کدام از متهمان به سه سال حبس و انتشار محکومیت قطعی متهمان از رسانه ملی، محرومیت از خدمات دولتی به میزان سه سال، دو سال محرومیت از انجام شغل مرتبط با حادثه‌ای که رخ داد، لغو پروانه مرتبط با شغل و محرومیت از عضویت در سازمان نظام مهندسی محکوم شدند؛ متهمان اصلی پرونده حسین عبدالباقی به میزان ۷۵ درصد بوده که به دلیل فوت او قرار منع تعقیب صادر شده و دیه از ناحیه اموال باقی مانده او پرداخت می‌شود.

Table 10

Persian News 5

Persian news 5¹

دادستان خوزستان:

رسیدگی و تحقیقات حادثه متروپل در شعبه ویژه با سرعت در حال انجام است.

آبادان - ایرنا - دادستان مرکز استان خوزستان گفت: رسیدگی و تحقیقات در خصوص حادثه متروپل آبادان با نظارت دادستان مرکز در شعبه ویژه با سرعت در حال رسیدگی است.

به گزارش جمعه شب ایرنا از روابط عمومی دادگستری آبادان، صادق جعفری چنگی، در جریان پیگیری آخرین وضعیت آواربرداری و مسائل حقوقی و قضایی حادثه متروپل آبادان اظهار کرد: با توجه به تاکید رئیس قوه قضاییه و دادستان کل کشور در برخورد با تمامی عوامل و مسببین حادثه روند رسیدگی

¹ <https://www.irna.ir/news>

و تحقیقات با نظارت دادستان مرکز در شعبه ویژه با سرعت در حال رسیدگی است.

وی ادامه داد: بخش زیادی از تحقیقات در خصوص حادثه متروپل انجام و با تکمیل تحقیقات در اسرع وقت تصمیم نهایی اتخاذ و پرونده به دادگاه کیفری ارسال خواهد شد.

دادستان مرکز استان خوزستان که به منظور پیگیری آخرین وضعیت آواربرداری و مسائل حقوقی و قضایی حادثه متروپل به آبادان سفر کرده بود، همچنین با تعدادی از خانواده‌های جانبازان حادثه ملاقات و خواسته و مطالباتشان را مبنی بر برخورد با تمامی عوامل موثر در حادثه استماع و به همدردی مجدد با خانواده‌های مصیبت دیده پرداخت.

تاکنون ۱۶ نفر از مقصران حادثه متروپل با صدور قرار تامین در بازداشت به سر می‌برند.

بخش‌هایی از برج دوقلوی در دست ساخت متروپل آبادان ظهر دوشنبه دوم خرداد ۱۴۰۱ دچار ریزش شد که در این حادثه ۴۳ نفر جان باختند و ۳۷ مصدوم شدند.

Persian Annotated Translation of Amparo Dávila's Tina Reyes Based on Vinay and Darbelnet's classification: a Translator's Comments

Mohammad Shokoohifar 

Master of Translation Studies, Translation Studies Department, Persian Literature and Foreign Languages
Faculty, Allameh Tabataba'i University, Tehran, Iran.
m.shoukohifar@gmail.com

Abstract

In this study, the text of Amparo Dávila's short story called Tina Reyes was used for translation and annotation. The purpose of choosing and translating this literary text was to sway the Persian-speaking reader and bring attention to the societal abnormality of rape. According to Halliday's register classification, the field and the tenor of the text were literary narrative and formal, respectively. Based on Reiss' classification, the text type was determined to be hybrid as it has two functions: expressive and operative. Maximum domestication and minimum censorship were applied according to the translator's ideology for translation. The translation strategy and translation procedures were chosen using Skopos theory and Vinay and Darbelnet's classification. The translation strategy was chosen to be target text-oriented. Modulation, transposition, and explicitation were the most widely employed procedures in translating process. It was concluded that if the translator keeps in mind the purpose and strategy of the translation until the end, the procedures are applied almost automatically and there is no need to make obsessive choices in finding equivalences. Schiavi's fingerprint theory was likewise validated at the end of the process.

Keywords: Annotation, Skopos, Genre, Strategy, Register

Citation: Shokoohifar, M. (2024). *Applications of Language Studies*, 2(2), 275-291.

1. Introduction

Although the study of annotated translation has been neglected in Iran and in English-Persian language pair, many articles and theses have examined the translators' problems and solutions adopted by them in other language pairs. For example, in English-Indonesian pair, Wibowo (2019) conducted a research study on the novel 'Any Minute' penned by Joyce Meyer, and Koskinen (2010) presented his thesis on the annotated translation of the Seventh Templar by Juha-Pekka Koskinen in Finish-English Pair. Scholars have used different models for annotation, but the purpose remains the same. In this paper, the researcher wants to see how and by which techniques the translators solve the problems based on the strategy they have in mind.

A 4840-word document was chosen for translation and annotation from English to Persian. Topics like the author of the text, rationale for translating this text, language role, register, and text type, will be discussed in the following sections. Then, depending on the readership and translation goal (Skopos), the relevant arguments for adopting translation methodologies are provided. Finally, techniques or procedures that have been used during the translation process, namely the annotations section, will be reviewed.

This story has been chosen from a collection of stories that have been published in the book entitled "Cuentos Reunidos" in 2009. The story is originally written in Spanish and then translated into English by Audrey Harris and Matthew Gleeson and published in the book entitled "The Houseguest and Other Stories" in 2018. Dávila takes the psychology of unrequited love and loneliness to its most excruciating extreme in 'Tina Reyes', possibly the greatest story in the book. Tina feels envious of one of her friends' apparently excellent status. Unlike her friend, who is happily married and her husband has an outstanding career, Tina does a job she despises, lives in a lousy neighborhood, and is lonely - so lonely that she believes it is her "destiny to be alone in the world". While she is on her way to see her friend, a stranger approaches her and attempts to converse with her. He starts off a little rough, but he does not do anything physically harmful. Tina is concerned by the man's presence, so she picks up her speed, almost running, till she reaches her friend's house safely. She meets the man again when she gets home, and he pursues her again, wanting only to chat, introduce

himself more wisely, beg for a bit of her time, and get to know her better. Tina becomes increasingly paranoid, believing that the man is planning to rape or murder her, and also that he definitely has a partner waiting in a nearby cab to take them both away. Dávila uses the presence of an ineffective guy to turn the screws on Tina over half of the whole narrative. He, too, is despondent and bereft. He starts to come across as a kind man. Tina's imagination, on the other hand, runs wild, trapping her in a world of danger.

As part of the Annotated Translation course, we were required to translate a text. Initially, I translated a legal text, but due to the difficulty in applying translation procedures like modulation, transposition, and others, and the need for greater accuracy, I switched to a literary text. From all the literary writings, I chose this one. I was deeply affected by the language of this story. It reminded me of the situation in Iran. Many Iranian ladies and mostly young girls are constantly confronted with these challenges. Because of the shattered trust between Iranian girls and boys, which is caused solely by the wrecked structures, I translated this text. The author translated this text to influence the readers and change their point of view about this anomaly. May all boys and girls be able to coexist peacefully in all societies.

2. Review of Literature

2.1. Language Role and Register

The field of discourse is "the total event, in which the text is functioning, together with the purposive activity of the speaker or writer; it thus includes the subject-matter as one element in it." Based on the classifications provided for discourse, the translated text falls into the realm of narrative. In general, it should be said that in this text, the tenor was quite formal. Even in conversations between the protagonist and his friend, we do not see very informal words. "The tenor refers to the type of role interaction, the set of relevant social relations, permanent and temporary, among the participants involved" (Halliday, 1994, p. 22). The following lexical items and sentences can be mentioned:

Lexical items: threshold, intertwine, frantic, sordid, whirlwind, enormous, spliced, thronged, vertiginously, unwound, collapse, voracious, devoured, harassing, confrontations, unleashed, boarded, crouched, engulfed, mercy,

invading, company, sinister, accomplice, deceiving, plunging, void, remarked, etc.

As an example of a relative complex structure, one can mention the following segment:

The confrontations, the two of them face-to-face and full of hate, she the target of everyone's gaze, the photographers harassing her, the medical examination, lying completely naked on a cold table, fastened by the wrists and ankles, and all of them above her like vultures, hands, eyes, on her, inside her, everywhere, and her naked in front of a hundred eyes that devoured her — never, never, she'd rather suffer whatever happened alone, in silence, without anyone else knowing.

2.2. Reiss's classification of texts

First, according to Katharina, we look at the type of the text. Reiss (1977, p. 108–109) proposes the key characteristics of each text type (genre) as follows:

1. *Informative*: information, knowledge, and views are examples of 'plain communication of facts'. The linguistic dimension utilized to transfer the information is logical or referential, and the content or 'topic' is the communication's primary focus.
2. *Expressive*: 'creative composition' refers to the author's use of aesthetic components of language. The author, or 'sender,' is highlighted, as is the message's format.
3. *Operative*: 'inducing behavioural responses': the operative function's goal is to convince or appeal to the text's reader or 'receiver' to do something, such as buying a product (if an advertisement) or agreeing to an argument (if a political speech or a barrister's final remark). The language is dialogic, and the emphasis is on appellatives.
4. *Audio-medial*: they are texts, such as films and visual and spoken ads that add visual imagery, music, and other elements to the other three functions.

According to Reiss's classification, at first glance, we may place the text in the second type, expressive function, because of its literary vocabulary and relatively complex sentences,

but there is also a critical view and encouragement of women and society to move. Criticizing the society can be clearly seen in many sentences. A few examples can be mentioned here:

Ex. 1: *"How terrible the police stations must be, the police themselves, the endless and degrading questions — what would he say?"*

In this part, we can see the author's criticism of the way police officers treat victims of rape and violence, which has been expressed quite openly.

Ex. 2: *"What would they think of her? Maybe that she'd been asking for it, they'd probably think she was "one of those girls," and they would treat her like one . . ."*

The pronoun "they", in this part of the text, refers to the general public, which refers to the kind of negative view of the victims of this social anomaly. Based on what has been discussed, the type of text cannot be classified as just "expressive," as we have seen criticizing and encouraging words in the text that aim to influence people's minds. The type of text we are discussing here is "hybrid," which has both an expressive and an operative role. This story is not presented with any visual content, so the fourth function — Audio-medial — is removed.

2.3. Readership

Unlike many scholars who have a prescriptive approach and consider good translation to be somewhat ST-oriented, in this translation, influencing the Persian-speaking reader was very important from the translator's point of view. Therefore, based on Skopos theory proposed by Reiss and Vermeer (1984/2013), the translation strategy was tried to be TT-oriented to impress the readers more.

At the beginning of the translation, it was decided to localize even the names of the characters, but this decision was abandoned due to the fact that some parts were not suitable for an Iranian context. Thus, if localization was done, the translator would have to cut those parts. Even an Iranian rape story could inspire the name chosen for the protagonist.

2.4. Translation Strategies and Procedures

At the beginning of the translation work, each translator decides with his or her own approach and mental "ideology" what "strategy" is appropriate for translating the text. Of course, there is also the fact that sometimes the translator translates only for profit, and according to the commissioner, he has to translate the text in a determined manner. But since in this case choosing the translation strategy was at the discretion of the translator, we examine the translation strategy and procedures according to Vinay and Darbelnet's classification (1995, 2004, p. 128–137).

In translating and annotating this text, the techniques presented by Vinay and Darbelnet have been used. They propose two types of strategies for translating texts namely, *direct* and *oblique* translation and each type has its own sub-types or procedures.

2.4.1. Direct translation

they consider this type as a literal translation. The authors prescribe literal translation for good translation: 'literalness should only be sacrificed because of structural and metalinguistic requirements and only after checking that the meaning is fully preserved' (Vinay & Darbelnet, 1995, p. 288). However, according to them, the translator may deem literal translation "unacceptable" for grammatical, syntactic, or pragmatic reasons. When a literal translation is not possible, they recommend using the oblique translation. However, after presenting their classification, the translator's justification for not following their prescription will be presented. Vinay & Darbelnet (1995) introduced and exemplified the following translational procedures in their seminal work when they originally compared English-French translations (in both directions) stylistically.

1. *Borrowing*: In pairs of languages where the letters of the alphabet are fundamentally different, such as Persian and English, this technique is accompanied by transliteration.
2. *Calque*: In this procedure, an SL term or structure is literally translated. The Persian calque ماه عسل is for the English honeymoon, for example.
3. *Literal Translation*: This is basically 'word-for-word' translation.

2.4.2. Oblique Translation

According to Vinay and Darbelnet, the translator may deem literal translation "unacceptable" for grammatical, syntactic, or pragmatic reasons. When a literal translation is not possible, they recommend using the oblique translation.

4. *Transposition*: This is when one part of speech is replaced with another.
5. *Modulation*: This procedure alters the SL semantics and point of view.
6. *Equivalence or Idiomatic Translation*: This phrase is proposed by Vinay and Darbelnet (1995, pp. 38–39; 2004, p. 134) to characterize situations in which languages utilize diverse stylistic or structural techniques to depict the same issue. When translating idioms and proverbs, equivalence comes in handy.
7. *Adaptation*: When a condition in the source culture does not exist in the target culture, this procedure involves modifying the cultural reference.

In addition to the above procedures, they introduce some other complementary techniques that are used in this translation, as follows:

1. *Amplification*: Sometimes because of syntactic expansion, the TL utilizes more words, e.g.: translating her position to جایگاه شغلی اش.
2. *Economy*: It is the opposite of amplification.
3. *Explicitation*: Implicit information in the ST is rendered explicit in the TT.
4. *Generalization*: The use of a more general word in the TT. For example: translating whispered to گفت.

Many scholars, including Vinay and Darbelnet, whose division was presented, have spoken about the strategy of translating. According to Skopos theory proposed by Reiss and Vermeer (1984, 2013), a translator can translate the text according to the purposes of translation. Although many scholars, such as Newmark, Vinay and Darbelnet have suggested source text-oriented translation for ease of work and speeding up the translation process, based on the Skopos theory, oblique translation or "communicative approach" as Newmark

proposed (1981, p. 39) is adopted to have a greater impact on Persian readers. Since the translator's ideology and aim for translating this story was to transfer the author's complaints to the setting of Iran and to bring out comparable anomalies as much as possible.

2.5. Ideology

Ideology is an indispensable part of any literary text as the writers' or translators' initial impetus for writing or translating that text arouses from the ideology they have in their minds. Ideology causes conscious manipulation in the text that besides the writers or translators' own ideology, it may be originated from the commissioner, editor, or institutional/governmental circles (Munday, 2016). In this work, as the translator's ideology was influencing the target readers and bolding the issue of rape in the target context, maximum domestication and minimum censorship were applied.

3. Method

Initially, the full text of the story was read once to determine its structure and content. Based on Skopos theory, the purpose of translation was determined. In the previous section, we discussed the purpose of the translation. As a result, the translator attempted to simplify the grammar and vocabulary as much as possible during the translation process. Contrary to Newmark's view (1981, p. 69) that the text should not be simplified too much and all the material "should not be placed on a plate and be handed to the reader", the translator did his best to ensure that the Persian-speaking reader could use the material from the "plate" with the least effort. Take his eyes like a spoon and eat them with his mind.

For translation, the translator tried to use the usual CAT tools, but since these tools, such as MemoQ and SDL Trados have errors and bugs in commenting in Persian, after several trials and errors, the translator finally decided to use Memsource which is a web-based CAT tool. Of course, this tool also has some problems with the Persian language, but it has a better performance compared to other ones.

The text of the story, like other stories and literary texts, especially some sentences and paragraphs, had its own difficulties. These were the specific parts in which the author

deliberately wrote the text in a more complex way, which naturally made the translation process more difficult.

The main problem was more in the field of Persian language, how the translator could reduce the effort of the Persian-speaking reader by distancing himself from the literal translation. In fact, it was somewhat difficult to follow the principle of equal impact that Nida (1964, p. 159) states.

The problems that the translator encountered in the translation process were commented on in the translation process. Of course, the problems that may have been somehow related to the translator's deficiencies, not knowing some words meaning and connotations, were invoiced from the commenting process.

A few days after the translation process was completed, the translator reread the entire text to correct and polish the translation and apply the necessary corrections. Then, by selecting the Vinay and Darbelnet 's categories that we discussed in the previous section, the translation techniques were listed in the Comment Section of the Bilingual Report, which was extracted from CAT tool.

4. Results

In this section, the frequency of each procedure used in the translation process as well as several examples of them will be presented.

Table 1

The Frequency of Translation Procedures

Procedures	Frequency
Borrowing	19
Calque	13
Literal Translation	15
Transposition	40
Modulation	110
Equivalence	26
Adaptation	7
Amplification	10
Economy	15
Explicitation	38
Generalization	2

some examples of commonly used procedures will be presented in order to have a better view of the procedures employed by the translator in dealing with the text.

Modulation:

ST: She was lucky enough to find a seat at this hour...

روز شانسش بود که آن موقع روز صندلی خالی پیدا کرد...

ST: If you thought about it...

اگر از این زاویه به ماجرا نگاه کنیم...

ST: ...trying to divert the course of her thoughts.

...سعی کرد به چیز دیگری فکر کند.

ST: She crossed a street at a red light and had to run to avoid being hit by a car.

چون چراغ راهنمایی سبز بود، مجبور بود بدود تا تصادف نکند.

ST: ...there was still a bit of daylight.

...هوا هنوز تاریک تاریک نشده بود.

Transposition:

ST: No one but her to do all the chores...

تمام کارهای خانه را تنهایی انجام می داد...

ST: They were always worried about money.

همیشه دغدغه مالی داشتند.

ST: She'd sleep peacefully hearing him breathe.

با شنیدن نفس های همسرش با آرامش می خوابید.

ST: ... sobbing in indignant protest until she fell asleep . . .

... و خشم و اعتراضش را با گریه و زاری ابراز کند.

ST: For a while Rosa went on joking with Tina about what had happened.

رزا چند دقیقه‌ای سر این موضوع تینا را دست می‌انداخت.

Explicitation:

ST: She was lucky enough to find a seat at this hour

روز شانسش بود که آن موقع روز صندلی خالی پیدا کرد.

ST: She didn't want to think about herself or her own life; it got her down — she always ended up depressed.

نمی‌خواست به خودش یا زندگی‌اش بیندیشد. وقتی که به این چیزها فکر می‌کرد، روح و روانش نابود می‌شد و همیشه چیزی جز ناراحتی برایش نداشت.

ST: Someone asked about a stop and the fare collector...

یکی از مسافران سراغ ایستگاهی را از کرایه‌چی گرفت...

ST: Against her will, Tina saw a woman with dyed blonde hair...

چشمش به صحنه‌ای افتاد که اصلاً دوست نداشت ببیند؛ زنی را با موهای بلوندشده دید که ...

Equivalence:

ST: ...butter wouldn't melt in their mouths,...

خودشان را به موش مردگی می‌زدند...

ST: Those words were like a lightning bolt suddenly striking her.

این جمله مانند تیری بر قلبش فرود آمد.

ST: She would come out alive...

آیا عمرش هنوز به دنیا است؟

ST: ...and her mind began to whirl like a spinning top.

و دنیا دور سرش می چرخید.

ST: Tina began to exhume her dead...

تینا از زنده و مرده اش برای او تعریف کرد

Translators face problems such as finding the appropriate vocabulary item or grammatical structure during the translation process. Since some of these problems were simply the result of not knowing the meaning of the words, they were removed from the annotation process. Here are some examples of translation problems and the translator's solutions mentioned during the translation process.

ST1: It would have been harder to watch them suffer for years and years, maybe it had been her destiny to be left alone in the world; she couldn't even keep a cat or a dog in that tiny room, and the poor canary Rosa had given her had died right away, doubtless for lack of fresh air and sunshine.

آن طوری برایش سخت تر می شد که سالها رنج کشیدنشان را تماشا می کرد؛ شاید سرنوشتش این بود که در دنیا تنها بماند. حتی نمی توانست گربه ای، سگی، چیزی را در آن اتاق کوچک نگه دارد. آن قناری بخت برگشته ای هم که رزا به او داده بود خیلی زود به خاطر کمبود هوای تازه و نور خوشید مرد...

Translator's comment:

معمولاً برای right away همیشه بلافاصله و بی درنگ به ذهن ما می آید. اما همش حس می کردم اینجا این واژه به درد نمی خورد. چون اگر حیوانی بلافاصله در شرایطی بمیرد، قطعاً انسان هم نمی تواند زنده بماند. پس به جای آن «خیلی زود» را گذاشتم که انگار بعد چند روزی مرده.

ST2: She would see countless couples come out singing or splitting their sides with laughter; sometimes they would get into fistfights there in the middle of the street, shouting the crudest insults at each other, then they would patch things up and disappear, arms around each other, down the dark streets; other times a patrol car would come and take them away.

زن و مردهای بی شماری را می دید که باهم بیرون می آمدند و آواز می خواندند یا با خنده از هم جدا می شدند. گاهی وسط خیابان یقه هم را می گرفتند و هر چه که از دهانشان بیرون می آمد نثار هم می کردند؛ کمی بعد، آشتی می کردند و دست روی شانه هم می گذاشتند و در آن خیابان های تاریک ناپدید می شدند. در غیر این صورت، ماشین گشت می آمد و جمعشان می کرد.

Translator's comment:

ترجمه هایی که برای عبارت arms around each other در اینترنت پیدا کردم دست در دست هم بود. اما حس می کردم این معنا را خوب منتقل نمی کند. مشکل ترجمه را با استفاده از تصاویر گوگل حل کردم.

ST3: Convinced of her fate, she meekly let herself be led along.

تقدیرش را پذیرفت و بی سروصدا همراه مرد رفت.

Translator's comment:

کلی برای meekly دنبال معادل گشتم ولی معادلی درست و حسابی پیدا نشد. متواضعانه، حلیمانه، با فروتنی، از روی افتادگی و... ولی هیچ کدام جالب نیستند برای اینجا. آخرش ساده ترین حالت ممکن برای توصیف این حالت را با استفاده از شم زبانی نوشتم یعنی بی سروصدا.

ST4: They had suffocated one poor girl with her own pillows, in her own house, after . . .

دختری بیچاره را در خانه خودش و با بالش خودش خفه کرده اند؛ اما قبل از آن...

Translator's comment:

تو این جمله خیلی سخت بود که چطوری after رو بیارم آخر جمله که اون سه نقطه هم که «سبک نویسنده است» بیاد آخر جمله. دینوتیشن مینینگ واژه نمی داشت که چیز دیگه ای به ذهنم بیاد. تا اینکه گفتم برعکس کنم و به جای بعد، بگم قبل. این طوری دیگه سه نقطه اومد ته خط.

5. Discussion

The procedures of oblique translation are used more than the procedures of direct translation in the text. Given the strategy that the translator had adopted for translating the text, the results are consistent with the theory proposed by Vinay and Darbelnet. Modulation, transposition, and explicitation had the highest frequencies among the procedures, respectively. Modulation is seen by Vinay and Darbelnet as "the touchstone of a skilled translator", whereas transposition "just demonstrates a very solid mastery of the target language" (2004, p. 246).

Even though it is difficult to trace the interactions that occur in the translators' minds until a solution is reached, the comments the translators mention about the text problems during the translation process can provide useful hints about the way they reach the solutions. The findings of this study can be used in teaching problem-solving skills in translation. This field has room for much research, so it is hoped that this research study will be the beginning of conducting annotation-based research on other texts with more details and a higher quality. As mentioned in the introduction, similar research has been done in other languages, but differences in translation in different languages require conducting separate research studies.

In the case of borrowing words, it is necessary to mention that most of these words are embedded in Persian and are used by general public, words like neon, rocking chair, blonde, drama, baseball, tone, and address. Since the narrative was initially written in Spanish and translated into English, and then translated into Persian, and because passive sentences are common in English literature, it is assumed that the frequency of passive sentences in Spanish is substantially lower than in English. They were rarely used in this story.

It can be even argued that the occurrence of passive sentences in Spanish is fewer than their occurrence in Persian, since when we examined the translation, we saw that the

passive sentences of the target text were written as active sentences in the source text. This is actually what we call "fingerprint" in Translation Studies.

Based on the results of this study, it can be concluded that the translators, who follow their strategy throughout the translation process, apply procedures related to the chosen strategy subconsciously] and do not need to be obsessively careful about the words and structures they choose. Also, the solutions to translation problems faced by translators are automatically influenced by the translation strategy. Another interesting and accidental result of this study was that no matter how much the translation is TT-oriented and the translator takes the reader's side, the "fingerprint" of the source language, the author, and the translator remains on the text (Schiavi, 1996, p. 14). Despite the widespread use of passive structures in English, the use of active structures in the Spanish text has also impacted the English text.

Conflict of interest

The author(s) certify/certifies that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in the present research paper.

References

- Dávila, A., Harris, A., & Gleeson, M. (2018). *The houseguest and other stories*. New York, New Directions.
- Halliday, M. A. K. (1994). *An Introduction to Functional Grammar*. London: Edward Arnold.
- Koskinen, V. (2010). *An annotated translation of passages from a historical novel by Juha-Pekka Koskinen* [M.Sc. Thesis]. University of Jyväskylä.
- Munday, J. (2016). *Introducing Translation Studies: Theories and Applications* (4th ed.). London: Routledge.
- Newmark, P. (2001). *Approaches to translation*. Shanghai: Shang Hai Wai Yu Jiao Yu Chu Ban She.
- Paepcke*, Munich Fink, 487-501.
- Reiss, K. (1977) 'Die literarische Übersetzung als Kommunikationsleistung', in K. H. Bender, K. Berger and M. Wandruszka (eds) *Imago Linguae. Beiträge zu Sprache, Deutung und Übersetzen. Festschrift zum 60. Geburtstag von Fritz*
- Reiss, K. (2000). Type, kind and individuality of text: Decision making in translation. *The translation studies reader*, 160-171.
- Reiss, K., Nord, C., & Vermeer, H. J. (2015). *Towards a general theory of translational action: Skopos theory explained*. Routledge.
- Reiss, Katharina & Vermeer, Hans J. (1984). *Grundlegung einer allgemeinen Translationstheorie*. Tübingen: Niemeyer.
- Schiavi, G. (1996). There is always a teller in a tale. *Target International Journal of Translation Studies*, 8(1), 1–21. Doi: <https://doi.org/10.1075/target.8.1.02sch>.
- Vinay, J. P., & Darbelnet, J. (1995). *Comparative stylistics of French and English*. Amsterdam: John Benjamins.

Wibowo, A. (2019). An annotated translation of passages from a historical novel by Juha-Pekka Koskinen. *Journal of English Education, Literature and Linguistics*, 2(1), 75–92. Doi: <https://doi.org/10.31540/jeell.v2i1.244>.

The Masculine Trauma in Katherine Mansfield's the Fly: the Boss' Traumatic Memories

Sahar Javadian Salemi 

Master of English Language Literature, faculty of Literature and Humanities, English Language Department,
Malayer University, Hamedan, Iran.
saharjavadian1998@gmail.com

Abstract

This study applies Sigmund Freud's psychoanalytic theory of trauma to Katherine Mansfield's (1922) work called 'the Fly'. It discusses the ways in which the protagonist of the story called the Boss attempts to deal with his traumatic memories. Freud in his *Studies on Hysteria* characterized the individual trauma as when a person is unable to react to a traumatic or affective memory in a way that successfully “discharges” those feelings, and therefore feels powerless in the situation. It is this feeling of powerlessness that leads to the trauma. This relationship between trauma and a lack of power manifests in the boss, which can be seen in his attempts to exert his hegemonic masculinity through the dominance and ownership at any given point of time. The idea of ownership is established throughout the story. The Boss wants to be able to feel traumatized at will or to have ownership on his own trauma so trauma becomes a part of the commodity economy in this particular narrative. It is concluded that although normally we associate trauma with loss and inability but trauma becomes the loss that has equitation of ownership.

Keywords: Trauma, The Boss, Masculinity, Repression, Freud

Citation: Javadian Salemi, S. (2024). *Applications of Language Studies*, 2(2), 292-306

1. Introduction

In Katherine Mansfield's short stories, the level of subtlety and psychological insight is striking. This research study makes an attempt to discuss Mansfield's art in creating the inner world of characters in *The Fly* (1922). Katherine Mansfield is a master in penetrating into the characters' psychological world. Her works concentrate on the inner mood, which is revealed to the reader vividly by her successful dealing with the story. The interior life like human isolation, disillusionment, paradoxes of idealism and reality, beauty and ugliness, joy and suffering are impressive in her stories. No matter what kind of character, innocent adolescent, man and woman in happy or unhappy marital life, old lady or man in his advanced age, Katherine Mansfield can show their complicated and subtle inmost feelings in depth.

First, instead of using strictly structured plots of her predecessors in the genre, Mansfield's stories are free from conventional plots (Kimber, 2015). Internal crisis has replaced external crisis. Mansfield skillfully diverts the readers' attention away from plot and invites them instead into the human mind. Second, when Mansfield shows her main character's feelings, she often uses her minor characters as a foil to her main character's psychological world. The minor characters either form distinct contrast with the protagonist or have close connection with the main character in terms of inner feelings. Third, Katherine Mansfield's peculiar gift also lies in her interpretation of subtle and unsayable feelings which lie hidden under men's everyday behavior and trivial events.

With much attention to small details of human behavior, Mansfield's stories of delicate plainness present exquisite inner feelings. Finally, the character's psychological world is presented to the reader also through the application of symbolism. Symbolism is constant and largely psychological in Mansfield's work. Abstract feelings are conveyed through numerous kinds of concrete symbols from material things to characters' habitual actions. In this way, Mansfield's stories offer an entrance to ordinary lives and subtle feelings (Kobler, 1990).

The word trauma comes from the Greek word meaning a physical wound, a corporeal wound, something which is inflated on the body. It is later used particularly in a medical and psychiatric literature. In Freud's stakes, the term trauma is understood as a wound inflicted

not upon the body but upon the mind. Freud believed that trauma was caused when a person experienced an event to which he/she was unable to react fully. In response, the individual attempts to repress the memory of these traumatic experiences; however, he or she is rarely successful. Instead, the memory resurfaces in the form of flashbacks against the individual's will or control and the individual is left seeking some kind of energetic reaction to the memory - some way to react in order to correct the traumatic memory. In "Beyond the Pleasure Principle" (1920), Freud describes a pattern of suffering that is inexplicably resistant in the lives of certain individuals. Whole idea of ontology and experience of trauma seems to suggest some degree of agency-less-ness one loses control, the motor control, the cognitive control, the experiential control and it extends into social political and other forms of control.

The backdrop of this particular story is the First World War. It is mostly about psychological violence of the war. Trauma becomes a very interesting category in the story. There is an equation which is established in this narrative between trauma and privilege. In other words, trauma becomes the ability by which one is able to feel traumatized and able to perform the trauma. This ability is connected to a sense of privilege, in the sense that one seems to have some ownership on his trauma. It is his own trauma which is more wounding and grieving compared to other people's trauma. So trauma becomes a category of agency, a marker of agency, and a marker of privilege which differentiates one particular traumatic person from the other people. The whole idea is about the hubris or the false pride, which an individual has in terms of being able to have ownership on his own trauma, in terms of having connected trauma with privilege, in terms of this equation between trauma and individuality. The protagonist of the story called the Boss is never named in the whole story, which is obviously quite symbolic because the whole idea of the boss is a degree of hegemonic masculinity about the character. So he wants to be the hegemonic male. He wants to be the dominant male establishing dominance. There is also the ability to feel traumatized at any given point of time.

The significance of this study is that I intend to explore how and in what ways Sigmund Freud's concept of trauma intersects with concept of masculinity in Katherine Mansfield "The Fly" (1922). In what follows, I will try to show that trauma in the boss can be

seen through his anxiety about insufficient masculinity and because of that he believes his son has died in the war, so he begins to test the masculinity of the fly. Additionally, his trauma is demonstrated by his efforts to suppress the emotions triggered by recollections of the traumatic event, as he strives to maintain his sense of masculinity despite feeling powerless in the face of the trauma. Both aspects of the story will be explored.

2. Review of Literature

Kaplan (1990) reads the fly in this story as meaninglessly killed by the boss, with the comment that this “unnerving portrait of victimisation, grief and suffering” might result from Mansfield’s identification of herself with a fly. Considering that Mansfield’s “The Fly” “incorporates a despair” Mansfield has experienced, Sylvia Berkman (1971) did not give a proper place to the fly.

In comparing “The Fly” with “Misery”, May (1994), argues that “The Fly” explores “the latent significance of the boss’s emotional state” (p. 202). He believes that this story emphasizes “the transitory nature of grief,” and his comment is restricted to only one character, the boss. His perspective, however, well captures the similarities between Mansfield and Chekhov, emphasizing on character as a mood, the “minimal dependence on the traditional plot,” and the “focus on a single situation in which everyday reality is broken up by a crisis” (p. 201).

Scholtmeijer (1993) observes that the nameless boss indicates his being merely as “one anonymous boss among others” who runs an undefined business typical in the city, while the nameless fly “holds steadier ownership of its individuality than the man, despite its depending for its life upon human caprice” (p. 165). However, her view—the fly, not being a symbol of “the power of life,” for its being finally drowned and thrown away, “extinguishes” “simultaneously the luxuriance in pathos and the hope of a moral cure”—appears to require reconsideration.

Kobler (1990) argues that the fly is “a replacement for [the boss’s] own wretched state” (p. 61). He believes that a parallel line between the boss’s dead son and the dead fly can be easily drawn. The son and the fly both are killed due to a pointless inhuman motive.

Though the boss has thought he could never recover from the loss of his son and told everybody that “[t]ime . . . could make no difference” (p. 359), he has overcome the death of his only son in less than six years, and so does he quickly forget the death of the fly, despite his meticulous attention to every detail of the fly’s movements, and even his consideration of “breathing on [the fly] to help the drying process” (p. 361). Like the old photograph of his son, “a grave-looking boy in uniform,” to which the boss does not want to “draw old Woodifield’s attention,” the existence of the fly appears out of place in the boss’s office where “the bright red carpet,” “the massive bookcase,” and “the table with legs like twisted treacle” give the boss “a feeling of deep, solid satisfaction”.

Harvey (2011) argues, interestingly, that the fly in this story “provides a *trompe l’oeil* effect just as it was used in fifteenth-century paintings” (p. 206); the fly makes this story truly real, by making this story as quotidian as possible, and, at the same time, by making us notice the existence of the fly in our human-dominated world. She believes that “The real mystery of this story is not what the fly signifies but what the fly is doing diving into the inkpot in the first place” (p. 207).

Parui (2017) highlights the boss’s refusal to confront his traumatic memory except as a means of ego preservation, portraying his rituals of mourning as a facade of strength and control. The boss's struggle to internalize and ritualize his loss is depicted as a constructed performance aimed at perpetuating his perceived manly authority. Interacting with a fly trapped in ink symbolizes the boss's futile efforts “to regain a sense of power and masculinity in the face of profound loss”. The narrative unfolds as a critique of masculine preservation projects and the inherent anxieties surrounding memory, mourning, and gender roles in the post-war era.

In the critical study of Boyle (1965), "The Fly" (1922) deals with the spiritual death of the boss and his replacement of material things with human things, as well as the similarities between Old Woodifield and the boss, and the parallels between the sufferings of the fly and those of the boss. The boss's encounter with the fly supports the idea that he is no longer spiritually alive and that his existence is just as sterile as old Woodifield's. The boss's troubles with his son's death are similar to the struggles the fly has with the ink blots. The boss had

anticipated being overtaken by sadness when Old Woodfield had described the boss's son in the first half of the story. However, at this point, he is unable to reply to the trigger for his sorrow. "My son," the boss moaned. The boss no longer reacts to the memories of his son, just as the fly no longer reacts to the blots of ink that land on him. The boss is so unperturbed by the thought of his son—a thought that used to cause him to experience sorrow spasms—that he is diverted by a fly in his inkwell. The boss is grabbed and startled by a "grinding feeling of wretchedness" after killing the fly.

3. Method

Freud is one of the earliest theorists of the mind. The psychologist who actually looked at art, literature, fiction, and stories in terms of having a more complex and fuller understanding of the mind and as something which is more complex in the brain. By his theories, literature becomes simply a complex representation of the mysterious mind. It does not become passive reflection of the human mind, rather; it becomes a representation. Both literature and psychoanalysis are engaging with the interplay of knowing and not knowing, between known and the unknown, and in doing. The idea of what remains unknown in actions and language becomes central to the ontology of trauma. Recognition of reality especially through the tragedy of trauma can be represented and not just reflected.

Freud also discussed the way individuals experience trauma, hysteria, and neurosis in much of his work including *Studies on Hysteria (1895)*, *Three Essays on Sexuality (1905)* and *'Civilized' Sexual Morality and Modern Nervous Illness (1908)*. For this article, we will be working with theories which were explained by Freud in *Studies on Hysteria (1895)*, and they were summarized by him in an early manuscript called "On the Theory of Hysterical Attacks". This is the key thing that one cannot leave behind the traumatic moment and it is always with oneself. It is internalized in a way that it almost becomes a part of one's flesh and blood, a degree of corporealization. We corporalize trauma and make that into part of one's body. The idea of traumatic neurosis can be described as an unwritten reenactment of an event that one cannot simply leave behind. Freud believed that the violence never goes away, the trauma never goes away, and that original moment of trauma is revisited again and again. That revisiting or reconnecting to the traumatic moment is exactly what is being

referred to by Freud as a traumatic repetition. It just takes one to the original experience of trauma and it is almost like a bodily experiential connection at that point where one touched the moment of trauma quite literally and that corporeal connection is exactly what makes us repetitive in quality.

According to Freud's theory, the way the experience of a trauma repeats itself exactly and unremittingly through the unknowing acts of the survivor and against his will would be to move away from trauma but then there is a certain degree of fixation which is a certain degree of being imprisoned to the traumatic moment. The subject seems to be stuck in some bubble of time where the traumatic moment keeps visiting him and all kinds of scenes and sounds around him remind him of the original traumatic encounter or the sight and movement of violence repeatedly.

As the author intends to explore the ways in which Sigmund Freud's concept of trauma intersects with the concept of masculinity in Katherine Mansfield's *The Fly*, here it has been tried to figure out how men work in this story. Our way of looking at masculinity studies in this particular story is to ask a very simple question, which is when a male character is working, how much of his actions or how much of his thoughts and his words are influenced consciously or unconsciously by his awareness of being a man that is to say, how much does he believe in what society has constructed as masculinity?

Masculinity studies are also very much predicated on a certain poststructuralist way of thinking which is to say that masculinity studies expose a binarity and then tell us that this binarity is untenable. It is not just one binary issue, there are several binary issues, on which masculinity is predicated. The most obvious binarity that we all recognize is male and female. There are so many other binarities. Therefore, when it is being said that this will separate the boys from the men, the boys therefore, are being attributed no masculinity. So the boys are not male, it is the men who are suitably male. So there is some kind of an adulthood that is automatically associated with masculinity. So the men are adults. And if one is not an adult, then he is a boy who is infantilized., there is a way in which the boy becomes infantilized in the same way that in a patriarchal society, women are infantilized. Thus, this concept will separate the boys from the men. It is basically saying that the boys have not acquired

masculinity yet. This is something that they must acquire. Having to go through trials or going through tribulations and that is when one acquires masculinity. Raewyn Connell is one of the critical figures in the field of masculinity studies. She describes four types of hegemonic masculinities for men: hegemonic, complicit, subordinated, and marginalized. Hegemonic masculinity describes the hierarchical interaction amongst multiple men and explains how some men make it look normal and that they do dominate most women and men. The second type of masculinity for men is complicit masculinity. As mentioned earlier, in a particular society, few men can live up to hegemonic masculinity ideals, but many support it. The concept of 'hegemony' normally suggests a large amount of consent. The other type of masculinity is subordinated masculinity. Masculinities encompass beliefs, values, behaviors and attitudes that fall outside the dominant meaning of what it implies to be man (Connell, 2005). It is in opposition to the hegemonic masculinity.

4. Results and Discussion

4.1. Performances of Masculinity in Characters

The term that Katherine Mansfield uses for the boss is "at the helm". If one is going to think about masculinity as comprising three levels, there are hegemonic masculinity, complicit masculinity, and input subordinate masculinity. We can certainly say that the boss as an embodiment of hegemonic masculinity is "at the helm", the way in which men should be at the helm. Woodifield belongs to the complicit masculinity. That is to say, he does not have masculinity himself, but he is at least supported because he says, "oh it is not in here". So this is the kind of masculine environment that Woodifield perhaps aspires to, but he knows that he will never have it. It is the boss who makes him aware of the fact that he is never going to have that, his masculine sense of authority like the upholstery, the furniture. Katherine Mansfield also reminds us that Woodifield is a frail, old figure (Mansfield, 2008, p. 359). So having to look at the way in which Katherine Mansfield, through her use of adjectives is creating Woodifield in the image of the inadequately masculine Woodifield is frail.

He is under the control of his wife and daughters clearly, everything that a true man should be. At the very beginning, there is a binarity that is drawn attention to because we do not know whether Katherine Mansfield believes in the binarity of the boy and the man. That

is certainly something that becomes very clear in the short story because the readers' feelings are important when Woodfield is described sitting in that chair opposite the desk from the boss. "The baby" is the word that the narrator used to describe what the Woodfield is. Very soon afterwards, the narrator tells us that he is under the control of his wife and daughter. So look at the way in which Woodfield is being constructed. As somebody who is inadequate in masculinity. Therefore, Woodfield is described as a baby and the readers are also told that he is basically under the thumb of his wife and daughters. What the audience gets to know about Woodfield is that he is a baby, so he has inadequate masculinity or he has no masculinity. Moreover, he is ruled by his wife and daughters, which means that this further exacerbates the lack of masculinity that he has. The whole hegemony of the boss is established by contrasting characters that are not hegemonic like old Woodfield. Although he is biologically younger than the boss, he is a faithful loyal servant, whose whole idea and old presence is establishing the boss's dominance.

Then, readers are introduced to another kind of masculinity. There is also another male spectrum, the dead male, boss's own son, who has been killed in the first world war, one whose loss was such a grievance for the boss. That morning he wants to perpetrate. In that sense, that becomes a very interesting equation between masculinity and vulnerability. Normally, we associate trauma with loss, with the inability but here it has the equations of ownership. That is why the whole idea of mobility and commodification becomes interesting in the story. But this is a very still masculinity because this masculinity is made available to us through a photograph of the son of the boss.

It is a staged photograph. It is very important to note that when the boss looks at the photograph, he says that, "my son wasn't like this" of course his son was not like this. It is very clear for us to figure out that the son of the boss was forced into a life that perhaps the young man did not want. Therefore, his son was a victim of the kind of pressure that society brings to bear on young men. In that photograph, we have the image of a victim of patriarchy. The boss goes on to show power and wealth and because he brings out that bottle of whiskey. He says that, "if you promise not to tell anyone, I'm going to tell you that this whiskey comes from the sellers of Windsor Castle" (Mansfield, 2008, p. 360). Windsor Castle, of course, is

the oldest castle that is used by the British Royal Family. So if one is talking about Windsor Castle, he is also talking about an extremely patriarchal kind of setup. The boss says that “this is where the whiskey comes from”. It can be understood that Woodfield is regarded as a very British name. The British royal family acquired the name of Woodfield only during the First World War. When King George the fifth realized that if they continued to retain the surname which was saxe-Coburg-Gotha, which was a German surname, then British people may actually start to suspect that their royal family has got German inclinations or that they are somehow favorably disposed towards the German, in order to create a sense of Britishness and consider themselves as being properly British and being not German but as British as Windsor. So there is a way in which patriotism was created by a very patriarchal man, George V. So when the boss says that the whiskey comes from Windsor Castle, Katherine Mansfield is also invoking a very patriarchal tradition of the way in which naming happens, and the way in which naming is used to consolidate patriarchy.

4.2. Masculine Trauma in “The Fly”

Freud believed that trauma was caused when a person experienced an event to which he/she is unable to react fully. In response, the individual attempts to repress the memory of these traumatic experiences; however she/he is rarely successful. Instead, the memory in the form of flash backs against the individual’s will or control and the individual is left seeking some kind of energetic reaction to the memory. This results in actions that the individual’s consciousness is unable to control or prevent. Trauma in the boss can be seen first through his anxiety about his insufficient masculinity making him believe that his son has died in the war, so he begins to test the masculinity of the fly. Second, his trauma is manifested by his struggle to control his emotions evoked after remembering the traumatic memory because nothing should weaken his masculinity even the trauma which we know it by feeling powerless in the situation. I will discuss the story in both ways.

4.3. The Death of the Fly and the Crisis of Masculinity

The Fly begins by introducing Mr. Woodified, an elderly retired man who, due to a stroke, is confined by his wife and daughters in a suburban home (ibid, p. 357). On Tuesdays, when he is briefly liberated from their care, he ventures to "the City" (ibid, p. 357) to visit his friend

known simply as "the boss" at his office. During the visit, Mr. Woodifield recounts his family's trip to the grave of his son Reggie and unexpectedly stumbles upon the grave of the boss's, who also perished in the war six years prior. After a casual conversation about family matters, Mr. Woodifield leaves the boss's office. The narrative then shifts to the boss preparing for a cathartic moment triggered by thoughts of his deceased son brought up by Mr. Woodifield. He instructs his employee Macey who has "dodged in and out of his cubby hole like a dog that expects to be taken for a run" (ibid, p. 359) to ensure no interruptions as he reminisces about his son, the future of his business, and his dashed hopes. However, as he anticipates a wave of sorrow, he realizes "he isn't as he wants to feel" (ibid, p. 360). In a perplexing turn of events, he notices a fly struggling in an inkpot and proceeds to inadvertently cause its demise by dropping ink on it out of curiosity. This unexpected distraction leaves the boss unable to recollect his previous thoughts.

There are different opinions on this horrible last part where the boss actually rescues the fly from the ink bottle and puts it on a blotting paper, allows the fly to get dry and then keeps on dropping little drops of ink till the fly can't survive that onslaught of ink drops anymore, by which time Woodifield has left. A parallel between the boss's dead son and the dead fly can be easily drawn (Scholtmeijer, 1993). The son and the fly both are killed due to a pointless inhuman motive. Though the boss has thought he could never recover from the loss of his son and told everybody that "[t]ime . . . could make no difference" (Mansfield, 2008, p. 359), he has overcome the death of his only son in less than six years, and so does he quickly forget the death of the fly, despite his meticulous attention to every detail of the fly's movements, and even his consideration of "breathing on [the fly] to help the drying process" (p. 361) (Kobler, 1990).

As said before I believe that the boss does the fly scenario because he is trying to test the fly to see how masculine the fly is. The fly of course is not interested in gender. The fly is not interested in proving his masculinity. The fly is simply interested in living. But the boss has to see that the fly succeeds in proving its masculinity. The boss is doing this because after the son's grave is mentioned, the boss is completely shattered inside. He is really troubled. He realizes that he has not mourned his son's passing in the war actually. So loss of the son is

something that the boss is troubled by. The part of that grief may be that the boss is obviously losing somebody that he loved but that is not the whole story. In losing the boy, the boss also lost that very patriarchal thing that somebody needs to continue his business, somebody who inherits the property. Therefore, the son was of value to the boss, not the boss is obviously losing somebody that he loved. The boss lost that very patriarchal thing, somebody to continue his business, somebody to inherit his property. The son not as the human being but the son was a value to the boss as somebody who is going to continue the boss's legacy. Thus, the boss loved his son for purely selfish reasons.

We are not very sure whether it was love at all. So, in dying, the boy seems to have actually tricked the boss through death. The son seems to have told the boss I will not be a part of your game, I will not take on your business, and this is something that I refuse to do. What is really happening over there is when the boss thinks that he wants to cry. He wanted, intended, and arranged to weep but of course he cannot weep. He cannot weep because of various reasons. First of all, he cannot because his grief is not entirely genuine. But the second reason why he cannot weep is because he has taught himself that emotional repression is a sign of masculinity. And now, he actually envies Woodifield because Woodifield, in spite of his being a baby, frail and old, is in touch with his emotions, which is something that the boss is incapable of doing. That is why when the boss actually starts to and really arranges to weep, he cannot, because the repression has become so solidified. It has become so deeply embedded that he cannot weep. And then of course, the fly episode happens and when the fly fails, the boss throws away the fly.

The throwing away of the fly is also the boss performing his contempt for a creature who has not succeeded in overcoming adversity, because the boss prides himself on overcoming adversity. What the boss perhaps realizes but then again, he capitulates, he returns back to his old self, he realizes that this overcoming of adversity has come to him at a great price. It has been a very expensive overcoming, he realizes that, but now of course it is too late. So instead of admitting to himself that in spite of his very posh office and his whiskey from the cellars of Windsor Castle, the boss is actually a failure as a human being. The boss throws away the fly, and by throwing away the fly, he reestablishes his masculinity

that had been cracked a little bit in the presence of Woodifield, not only in the presence of Woodifield, but when Woodifield has mentioned his son's grave. Therefore, when he forces those tears, the tears do not come. After the fly has been thrown away, the boss tries to remember what was it that he was intending to cry about but the story ends when he could not remember.

4.4. Horror of Loss of Beloved Ones

Everything in the story is also very complex, because normally as we know, the whole idea of hysteria is seen as a female malady. The very sexist definition of hysteria, as a female malady sometimes happens all the time to a woman, a woman who is more hysteric naturally. But hysteria becomes a performative category in the story that one is able to feel hysteric for this story because he is so manly. Thus, it becomes an interesting equation between masculinity and hysteria. So one's masculinity privileged masculinity and hegemonic masculinity is determined by the ability to be hysteric. Thus, hysteria becomes ability. The boss in the story wants hysteria to become a performative category. He wants to perform hysteria at will; he wants to be able to feel traumatized at will (Parui, 2017). In this way as discussed earlier, therein lies the privilege masculinity that he wants to embody. So his privilege as a male, the fact that he is a privileged male is determined by his ability to be hysteric at any given point in time. His privilege as a male is determined by his ability to feel traumatized at any given point in time. That really makes the story very complex and bizarre at many levels. The strangeness in the story is about how trauma becomes a male category and male commodity to some extent that one is able to acquire trauma, able to have ownership on his own trauma, and his property. That is why there is commodification of trauma and hysteria, when it becomes a commodity owned by a man The boss is sitting in his own office, he owns a lovely office. He owns all the gadgets around him. He owns the capital that flows inside the office. Therefore, in that sense, trauma also becomes part of the commodity economy in this particular story. So he wants the ability of going back to the original hysteric traumatic moment and wants to replay it very professionally.

5. Conclusion

In her story “The Fly”, Katherine Mansfield (1922) has chosen a figure as a representative of conventional masculinity: the boss is authoritative, gross, boorish, callous, combative, competitive, proud and self-complacent. He needs to feel that he must have a control over all his affairs, even superiority and power over two inevitably uncontrolled matters; the mortal condition and emotional responses to the past traumatic memory. In spite of the dislike the readers develop towards him, they also feel sympathetic to the boss's existence who is torn apart because of the strict rules of hegemonic masculinity.

In this story, Katherine Mansfield implemented the boss to show the readers this is what men become if they believe in masculinity so much. They become so emotionally stunted. They become so emotionally repressed that even when that repressed emotion threatens to bubble up, they push it down. After pushing it down, they tell themselves that they cannot remember what had just bubbled up and what had just been pushed down. They even forget that they are repressed. So a part of being masculine is to forget your own emotional repression. Mansfield is taking a very critical distance in this story to look at masculinity in this incredibly remarkable way.

References

- Berkman, S. (1971). *Katherine Mansfield: A Critical Study*. Archon Books.
- Boyle, E. (1965). *The Death of the Boss: Another Look at Katherine Mansfield's "The Fly"*. *Modern Fiction Studies*, 11(2), 183-85. Doi: <https://www.jstor.org/stable/26278801>.
- Connell, R. W. (2005). Imperialism, and masculinities. *Handbook of studies on men and masculinities*, 71.
- Freud, S., & Breuer, J. (1895). Studies on Hysteria. In *Standard Edition of the Complete Psychological Works of Sigmund Freud* by James Strachey (1955), Hogarth Press.
- Harvey, M. (2011). "Katherine Mansfield's Menagerie." In *Katherine Mansfield & Literary Modernism* by Janet Wilson, Gerri Kimber & Susan Reid. London: Continuum, 202-210.
- Kaplan, S. J. (2010). *Circulating Genius: John Middleton Murry, Katherine Mansfield and DH Lawrence*. Edinburgh University Press.
- Kimber, G. (2015). *Katherine Mansfield and The Art of the Short Story*. Palgrave Macmillan.
- May, T. (1994). *The political philosophy of poststructuralist anarchism*. Penn State Press.
- Kobler, J. F. (1990). *Katherine Mansfield: A Study of the Short Fiction*. Boston: Twayne Publishers.
- Mansfield, K. (2008). *Selected Stories*. Oxford: Oxford University Press..
- Parui, A. (2016). For the life of him he could not remember': Post-war Memory, Mourning and Masculinity Crisis in Katherine Mansfield's 'The Fly'. *W. Martin (Author) & C. Hanson & G. Kimber (Eds.), Katherine Mansfield and Psychology*, 113-124.
- Scholtmeijer, M. (1993). *Animal victims in modern fiction: From sanctity to sacrifice*. University of Toronto Press.