In the Name of God

Scientefic Quarterly Journal of
Language Horizons
Volume 7, Issue 1, Spring 2023 (Serial No. 15)

Publisher: **Alzahra University**
Managing Editor: **Dr. M. Shirazizadeh**
Editor-in-Chief: **Dr. M. R. Anani Sarab**
English Editor: **Dr. Sh. Teimournezhad**
Persian Editor: **Dr. N. Monfared**
Managing Director: **N. Jafari**

Page Layout, Printing and Binding**: Mehrravash Pub**. (+98 91 99 19 92 42)

Alzahra University Publication
Address: Alzahra University, Vanak, Tehran, Iran. Postal Code: 1993891176
Email: lghorizons@alzahra.ac.ir
ISSN**: 2588-350X** E-ISSN**: 2588-5634**

The Journal of Language Horizons is published by **the Iranian Linguistic** Association

the Iranian Linguistic Association

Journal of Language Horizons

According to the Ministry of Science, Research and Technology proclamation dated 1397/12/06 and numbered /3/18/311330, the *Journal of Language Horizons* was granted full Academic-Research status by the ministry's publication committee in their meeting of Jan. 30, 2019. This status has been applied from Volume 1, Issue 1, Spring/Summer 2017.

It is a monolingual (English) journal which operates on a blind peer review policy; it publishes twice- yearly (biannually) papers which report the findings of original research on the current trends and topics in different language related issues.

Submission of manuscripts should be sent electronically via Journal website:
http://lghor.alzahra.ac.ir.

More information about guidelines for submission and the policy of the Journal is provided in other parts of the website. Only articles received online through the site will be sent to referees.

Office Address: Journal of Language Horizons, 2nd Floor, Faculty of Literature and Languages, Alzahra University, Tehran, Iran

E-mail: lghorizons@alzahra.ac.ir

Aims and Scope

Considering the National Master Plan for Science and Education in the country and the necessity of *localizing* the humanities, this journal welcomes English articles contributing to this aim.
Aims of this journal include but are not limited to:

- Promoting scientific and professional knowledge of researchers in the below mentioned areas.
- Sharing the findings of researchers in the mentioned fields.
- Promoting general and professional knowledge necessary for learning English as an international language among Iranians and international interaction in the future.
- Promoting Iranian culture along with the English.

The main focus of the Journal is on research conducted on language learning and teaching. We are in particular interested in research papers on L2 education (in particular EFL/ESL), in a variety of levels, including school level, university level, institute level, etc. The Journal welcomes papers on teaching and learning any component of language including skills and sub-skills, as well as sociolinguistics, psycholinguistics, discourse and pragmatics, TESL or TEFL, Teaching Persian to Speakers of Other Languages (AZFA), language evaluation/testing, language and culture, teaching language for specific purposes, teaching methodology, English and Persian literature, language studies, translation, and linguistics but written only in English.

Submission Guidelines

Authors are encouraged to submit complete unpublished and original works which have not already been published elsewhere nor are under review in any other journal. They are kindly requested to refer to the Journal website: http://lghor.alzahra.ac.ir/ to register according to the instructions given. Once registered, each author will be provided with a username and a password enabling him/her to follow up the submission process. The next step is to send the paper to the Journal to be considered for review.

The manuscripts submitted to *Journal of Language Horizons* should follow the latest version of APA style. Contributing authors are advised to read this document carefully and adhere to the instructions given below before sending their papers to the Journal office. The article will be sent in two separate files including the title file and the manuscript file. The title file includes the title, author's

/authors' name(s), their institutional affiliation(s), phone number(s), and the email(s), preferably the academic one(s). In case of multiple authors, the corresponding author should be indicated.

The second file, the main manuscript, starts with the abstract. The abstract should be between 150 and 250 words, followed by five keywords, separated by commas. The abstract should include information on the purpose of the research and/or research question(s), the methods and materials used, information on the analysis procedures as well as the major findings. The Persian translation of the abstract should be provided too. The font type should be Times New Roman and the size is 12. The whole manuscript should be double-spaced throughout and the new paragraphs should be indented. The manuscript should be divided into clear sections such as: Introduction, Review of literature (which may include sub-sections), Method (including participants, materials, and procedure), Results, Discussion, Acknowledgements (if necessary) and References (and Appendices, if needed). The reference list should be on a new page, double spaced, and the hanging indent method should be used (all lines after the first one are indented). The length of the paper should be between 7000 and 10000 words. The following pages provide key information and give examples of APA style. More information on APA can also be found in the website: http://www.apastyle.org/

REFERENCES

All in-text citations must be listed in full in the reference list at the end of each article following the specifications of the *APA* manual and all references listed must be cited somewhere in the text. Begin the reference list on a separate page entitled "References" and double-space it throughout. Each entry must include the author's name, co-authors (if any), publication date, and title of work. For a journal article, also provide the name of the journal, volume and issue numbers, and page numbers for the article. For an article in an edited volume, list the editor's name, title of the collection, and page numbers of the article. For a book or monograph, list the edition, volume number, series, and name of publisher. For all online works, the DOI should be provided as well. If not available, the URL can be provided instead. Punctuate and capitalize as in the following examples:

Book

Brown, G., & Yule, G. (1989). *Discourse analysis*. Cambridge University Press.

Journal article

Bacon, S. M., & Finnemann, M. D. (1990). A study of the attitude, motives, and strategies of university foreign language students and their disposition to authentic oral and written input. *The Modern Language Journal*, *74,* 459-473.

Journal article with volume number

Lee, M-B. (2002). A closer look at language learning strategies and EFL performance. *Foreign Languages Education*, *10*(1), 115-132.

Article in a book

Dechert, H. (1983). How a story is done in a second language. In C. Faerch & G. Kasper (Eds.), *Strategies in inter-language communication* (pp. 175-195). Longman.

Journal Article with DOI

Paivio, A. (1975). Perceptual comparisons through the mind's eye. *Memory & Cognition*, *3*, 635-647. http://doi.org//10.1037/0278-6133.24.2.225

Journal Article without DOI (when DOI is not available)

Hamfi, A. G. (1981). The funny nature of dogs. *E-journal of Applied Psychology, 2*(2), 38-48. http://www.ojs.lib.swin.edu.au/index.php/fdo

Encyclopedia Articles

Brislin, R. W. (1984). Cross-cultural psychology. In R. J. Corsini (Ed.), *Encyclopedia of psychology* (Vol. 1, pp. 319-327). Wiley.

Developmental genetics. (2005). *In Cambridge encyclopedia of child development*. http://www.credoreference.com.library.muhlenberg.edu:80/entry/cupchilddev/developmental_genetics

Reference citation in text

Direct citation 1
She states, "the 'placebo effect' ... disappeared when behaviors were studied in this manner" (Miele, 1993, p. 276), but she did not clarify which behaviors were studied.

Direct citation 2: Fewer than forty words
Lee (1999) found that "The EAP writing curriculum incorporate reading and analysis of major academic journal articles in the specific field to identify macro-level organization an obligatory 'moves' in conjunction with writing practice." (p. 21)

Direct citation 3: More than forty words
Miele (1993) found the following:
The "placebo effect," which had been verified in previous studies, disappeared when behaviors were studied in this manner. Furthermore, the behaviors were never exhibited again, even when reed rings were administered. Earlier studies were clearly premature in attributing the results to a placebo effect. (p. 276)

One work by two authors
- Smith and Takamoto (1997) argued that. ...
- In recent study of SLA (Smith & Takamoto, 1997) ...

One work by multiple authors: Use et al.
- First citation: Jones et al. (1997) found that...
- in a recent study of second language acquisition (Jones et al, 1997)....
- Subsequent citations: Jones et al. (1997) further argued that. ...
- In a recent study of second language acquisition (Jones et al., 1997)...

One work by multiple authors, more than 6
- First citation: Dresler et al. (1992) showed that...
- In a recent study of second language acquisition (Dresler et al., 1992)...

Multiple works: Arrange in alphabetical order
Previous research (Lass, 1992; Meyer & Sage, 1978, 1980; Nichols, 1987a, 1987b; Oats et al., 1973)...
For works not included in the above examples refer to *Publication Manual of the American Psychological Association* (7th ed.).

Review Policy

Since the Journal uses a double-blind review system, it is essential that all author identifying information be removed from the paper and that author(s) information should only be provided in the title page. In case the author's/authors' work is mentioned in the manuscript, replace the name(s) with *Author(s)* in the the text. In other words, there should be no trace of the author in the text.

Before submitting your article to be considered for publication in the Journal of *Language Horizons*, make sure that it adheres to all guidelines given in this document; otherwise, the Journal office may reject the paper before sending it out for review. The papers following these general criteria may be submitted to be evaluated for possible publication. Please be informed that the Journal takes plagiarism very seriously and the contributors are cautioned against this.

The journal follows a rigorous reviewing policy. Each submitted paper is first evaluated for its style consistency and appropriateness of the topic. If found faulty or not appropriate, the paper is returned to the corresponding author for further work and resubmission. The papers that meet initial submission criteria are then reviewed by members of editorial and advisory board as well as external review-ers. A final decision is made on the status of the paper based on the feedback offered by board members and anonymous reviewers to the Editor in Chief. The final decision will be in form of 1) Accepted as it is; 2) Minor revisions; 3) Major revisions; and 4) Rejected. The Journal keeps the right of literary and technical changes but not changes in the content of the articles. Also, article publishing depends on the editorial board's approval.

Scientefic Quarterly Journal of Language Horizons, ***Alzahra University***
Volume 7, Issue 1, Spring 2023 (Serial No. 15)

Contents

Journal of Language Horizons

Scientefic Quarterly Journal of Language Horizons, *Alzahra University*
Volume 7, Issue 1, Spring 2023 (Serial No. 15)

An Investigation of the Relationship Between Subjective and Objective Cognitive Load Measures of Language Item Difficulty

Research Article
pp. 7-33

Shadi Noroozi[1]
Hossein Karami*[2]
Zari Saeedi[3]

Received: 2021/10/01 Accepted: 2022/03/16

Abstract

The current study strived to delve into the response behavior and perceptions of examinees while taking a test in light of cognitive load theory. The empirical data were collected from 60 MA English major graduates and students, with a high level of language proficiency. The participants were required to answer 60 multiple-choice language items (i.e., grammar and vocabulary questions), taken from the high-stakes tests of the MA English majors of the Iranian university entrance examination (IUEE), as fast and as accurately as possible. After completing each test item, they rated their perceptions with regard to the difficulty of test items (Bratfisch et al., 1972) and the amount of mental effort (Paas, 1992). Their response time spent on each language item and their selected options were also stored by the Psychopy software (Peirce et al., 2019). Through running Pearson and Spearman rho correlations, the findings revealed that response time enjoyed a strong positive correlation with mental effort, meaning that both objective and subjective cognitive load measures matched in terms of

* Corresponding Author
[1] MA Holder, English Department, Faculty of Foreign Languages and Literature, University of Tehran, Tehran, Iran. shadi.noroozi@ut.ac.ir
[2] Assistant Professor, English Department, Faculty of Foreign Languages and Literature, University of Tehran, Tehran, Iran. hkarami@ut.ac.ir
[3] Associate Professor, Department of English Language and Literature, Faculty of Persian Literature and Foreign Languages, Allameh Tabataba'i University, Tehran, Iran. saeedi.za@atu.ac.ir

DOI: 10.22051/lghor.2022.37418.1551
DOR: 20.1001.1.2588350.2023.7.1.1.9

sensitivity to cognitive load changes in language test items. Further, the subjective measures of perceived mental effort and perceived level of difficulty revealed to be the sound indicators of cognitive load changes. As predicted, response time also indicated that more difficult language test items imposed a greater amount of load. The implications of the study will be explained.

Keywords: cognitive load, language test-items, multiple-choice questions, response-time, subjective/objective measures

Introduction

Test developers and psychometricians have voiced their concern over the understanding of and improving the psychometric qualities of language tests in recent years. The remarkable effects of high-stakes tests not only on the individuals' academic career (Shohamy et al., 1996) but also on their cognitive architecture or minds (Sweller et al.,1998) have recently received considerable critical attention. The exploration of cognitive processing provoked by test items is becoming fundamental to the examination of language test items because cognitive processes can have undue influences on the individuals' language learning and their performance on tests (Gass et al., 2013; Ponce et al., 2020). In this regard, the effect of tasks (items) on individuals' minds has been investigated with respect to cognitive processing using cognitive load theory (CLT) (Dindar et al., 2015).

Cognitive load theory was presented by Sweller in the 1980s and this theory can account for cognitive load patterns as to multimedia learning (Brünken et al., 2003; Paas et al., 2008; Whelan, 2007; Wiebe et al., 2010), instructional materials, learning, and teaching (Sweller et al., 1998, 2019). Concerning language testing, cognitive load can be expressed as the distribution of cognitive capacity of test takers while taking the test (Sweller, 1988). Hence, cognitive load plays a pivotal role in language test performance. Of the three classifications of cognitive load theory (i.e., intrinsic, extraneous, and germane load), intrinsic load corresponds to inherent tasks (items) characteristics (de Jong, 2010) which can be related to the difficulty of tasks (Paas et al., 2003). Substantially, CLT can be regarded as a new line of research on the explanation for the difficulty of some material in comparison with some other material (Martin, 2014).

Subjective (e.g., perceived mental effort and difficulty level), objective

behavioral (e.g., response time and reaction time), and physiological (e.g., electroencephalography, pupil dilation, heart rate, etc.) measures are prevalent in cognitive load studies. Mental effort appears to have a relationship with the processes or the cognitive capacities dedicated to accomplishing a task, whereas perceived level of difficulty seems to be related to the difficulty of a task or item itself (van Gog & Paas, 2008). On the other hand, it has been shown that several measures should be used to paint a complete picture of cognitive load (Leppink, 2017; Skulmowski & Rey, 2017). Thus, considering this issue can provide insights into whether there is a correspondence between cognitive load measures using tasks (items) with varying degrees of difficulty. Consequently, the current study set out to scrutinize the behaviors and perceptions of test takers while taking the multiple choice vocabulary and grammar questions of IUEE in light of CLT.

Literature Review

Cognitive Load Theory (CLT)

This theory centers on the "unobservable" phenomena (i.e., cognitive load) that individuals experience while dealing with different tasks. Central to the entire discipline of cognitive load theory (CLT) are the constructs of *cognitive load* and *learning*. This theory has long been an object of research in a wide range of scientific domains such as cognitive psychology and instructional design (Sweller, 2010). According to Sweller et al. (1998), CLT is a framework that explains the relationships among learning, cognitive architecture, and materials design. It is pivotal to know about the cognitive architecture and the efficiency of instructional designs in CLT.

Cognitive load theory rests on the assumption that cognitive architecture of human beings consists of two sets of memory stores: limited working memory and unlimited storage memory (Martin, 2014). As one of the elements of cognitive architecture, working memory, a place for the occurrence of all conscious cognitive processing, is responsible for information processing (Paas, 1992). Nonetheless, this type of memory is restricted in capacity to hold and process information (Miller, 1956; Peterson & Peterson, 1959). Due to this limitation, simultaneous processing and connecting several elements of a task may considerably surpass working memory capacity. Hence, learning may be hampered (Chandler & Sweller, 1991;

Paas, 1992).

In fact, the imposed load on the working memory is directly affected by task requirements (Sweller et al., 1998). To avoid the unnecessary cognitive load or overload imposed by a task, working memory capacity limitations should be considered in instructional designs (Sweller et al., 2011). Cognitive overload and underload may have adverse effects. As for cognitive overload, Johannsen (1979) reasoned that too much amount of load can negatively affect the functioning of working memory. On the other hand, cognitive underload can influence the performance on a task due to the lack of motivation, for example (Young et al., 2015).

To have effective learning, instructional designs should be reconsidered and modified by reducing working memory load. In other words, it seems necessary that the instructional design be proportional to cognitive architecture (Schnotz & Kürschner, 2007). In addition to learning, performance is also affected by the overload or underload. As an aspect of CLT, *performance* is concerned with the correct responses, error rate, and response time (Paas et al., 2003). Poor performance can be ascribed to task demands surpassing cognitive capacity. Stated more specifically, the demands of a task can influence the amount of load imposed on the minds of the test takers and hence their performance (Dindar et al., 2015; Gvozdenko & Chambers, 2007). Therefore, test takers' performance can indicate the load imposed by the test items.

Intrinsic Load

Due to the interaction between instructional design and cognitive architecture, several forms of loads (i.e., intrinsic, extraneous, and germane) have been identified (Paas et al., 2003; Sweller, 1994; Sweller et al., 2011; Sweller et al., 1998, 2019). Intrinsic load may be defined in terms of the innate nature of a task (Sweller et al., 1998). Central to intrinsic load is *element interactivity* which deals with the simultaneous processing of several elements in the working memory to understand or learn a task (Sweller, 2010). Therefore, the relationship between working memory capacity and element interactivity becomes paramount. Pollock et al. (2002) recognized grammatical syntax as an instance of high interactivity material. In addition, there is a positive relationship among element interactivity, intrinsic load, and working memory capacity (Sweller, 2010).

For de Jong (2010), intrinsic load can refer to the experienced difficulty of a task. In other words, the difficulty of materials can be associated with their intrinsic or inherent nature. According to a definition provided by Sweller (2020), the term *difficulty* may be defined with respect to the nature of the information of a task and the individual's knowledge. Difficulty may have different sources (Sweller et al., 2011). Indeed, the degree to which task elements interact is related to intrinsic load. Regarding a task with low element interactivity, the simultaneous processing of a great number of elements seems to be the cause of difficulty in learning the vocabulary of a language (e.g., learning the translation of "dog" into Persian).

Measuring Cognitive Load

Traditionally, measuring cognitive load was restricted to error rate. As the theory developed, more direct measures of cognitive load burgeoned (Sweller et al., 2011). Due to the lack of any single standardized method (Brünken et al., 2010), the implementation of diverse measures is pivotal to obtain a more precise picture of cognitive load (Leppink, 2017; Skulmowski & Rey, 2017). The validity and reliability of measures are of great concern to assess the load imposed by the instructional designs and experienced by the individuals (Brünken et al., 2010).

Regarding the subjective measures, self-report questionnaires are used as for rating perceived mental effort and level of difficulty. Learners can be asked to reflect on their invested mental effort (Paas, 1992). Brünken et al. (2010) maintained that behavioral parameters were taken as the indicators of cognitive load. Neuro-imaging techniques, time-on-task, the secondary task, and reaction time are different methods identified to measure cognitive load objectively.

Cognitive Load Theory: Review of Empirical Studies

Concerning the implementation of diverse instruments, there is a wave of studies that applied cognitive load measures to the fields of multimedia learning (DeLeeuw & Mayer, 2008; Dindar et al., 2015), task-based language teaching (TBLT) (Lee, 2019; Révész et al., 2016; Sasayama, 2016), and testing (Pouw et al., 2016; Prisacari & Danielson, 2017).

To investigate the underlying trait of cognitive load, DeLeeuw and Mayer (2008) carried out two experiments through subjective and objective measures in

multimedia learning. They examined the sensitivity of measures through the use of different levels of sentence complexity, problem solving situations, and redundancy program. As for the sentence complexity, the high complexity sentence contained more interacting elements in comparison with the low complexity sentence. Sentence complexity can potentially cause intrinsic load (Sweller, 1999); thus, the higher the element interactivity of a sentence, the more complex the sentence and hence the greater the intrinsic load. Besides, reaction time, a meaningful indicator of cognitive load, showed that the longer the reaction time, the higher the cognitive load.

In their first experiment, they found a significant positive correlation between mental effort and sentence complexity: the higher the sentence complexity, the greater the amount of invested mental effort. A positive correlation was also found between reaction time and high complexity sentence. Moreover, the study revealed a modest significant correlation between reaction time and mental effort. Similar to the first experiment's results, a higher amount of mental effort was observed in the case of high complexity sentence in their second experiment. In contrast to the first experiment's findings, reaction time did not measure precisely task difficulty with regard to the number of interacting elements. In addition, no significant correlation was found between mental effort and reaction time.

In one study undertaken by Dindar et al. (2015), the difference between the cognitive loads of two different types (static vs. graphic) of achievement tests was investigated by making use of response time, rate of accuracy, subjective measure of mental effort, and the secondary task. The results demonstrated that response time was a reliable index of cognitive load: the longer the response time, the more complex the task, and the greater the load. Previous studies reported a modest correlation between reaction time and mental effort (DeLeeuw & Mayer, 2008; van Gerven et al., 2006); however, no statistically significant correlation was observed between the variables.

As for the context of TBLT, numerous studies have been designed to examine the validity of task complexity with regard to subjective and objective measures widely implemented in cognitive load studies (e.g., Lee, 2019; Révész et al., 2016; Révész et al., 2014; Sasayama, 2016). Sasayama (2016) conducted research to assess the cognitive complexity of the narration of four picture sequences

through three measures: time estimation, self-rating, and the secondary task. After each narration, perceived level of difficulty and mental effort in addition to time estimation were measured. The four tasks had varying difficulty, ranging from the simplest to the most complex. The participants were also instructed to respond quickly to the secondary task. The findings revealed that the most complex task required the longest reaction time. Moreover, the most complex task was considered the most difficult and seemed to require the greatest amount of mental effort. Sasayama (2016) highlighted the impact of proficiency not only on the participants' performance on the same task but also on the cognitive load measures. In other words, proficiency seemed to affect the performance of learners and cognitive load measures differentially. Her findings also suggested that the reported effects on the cognitive task complexity were over-estimated through the use of self-reports. Therefore, it was advised that the interpretation of the findings obtained through the subjective measures be made with caution.

In addition, Révész et al. (2016), using the measures of self-ratings, expert judgments, and reaction time, investigated the validity of task complexity. Their results indicated that the more complex the task, the greater the amount of consumed mental effort. The findings also revealed that more complex tasks were identified as more difficult. No statistically significant difference was found between the different primary tasks' reaction time and their different levels of complexity, meaning that reaction time had no relationship with task difficulty.

In another study, Lee (2019) examined whether variations in task complexity could truly lead to alterations in cognitive load by the implementation of the self-ratings of perceived mental effort, stress, difficulty, and time estimation, and the implementation of the objective measure of the secondary task. The results showed that the most complex tasks were perceived to be the most difficult. Furthermore, these tasks enjoyed the highest rate of mental effort. The results indicated a positive relationship between response time and task complexity. In contrast to the earlier studies that highlighted the importance of accuracy rates in the case of more complex tasks (Révész et al., 2016; Révész et al., 2014), Lee (2019) observed significant effects for the amount of time rather than accuracy. Similarly, Sasayama (2016) reported no effects for accuracy. In other words, reaction time was the longest regarding the most complex tasks compared to the ones with the least

complexity.

On the other hand, Lee (2014) examined the validity and reliability of cognitive load measures through electroencephalography (EEG), self-ratings, and learning outcomes. After watching a seven-minute documentary video, the participants were directed to report quickly their perceived mental effort and difficulty. The findings revealed a statistically negative correlation between difficulty ratings and learning outcomes. In other words, as the task became more complex and imposed a greater amount of load, it was perceived to be more difficult. The participants appeared to cease expending mental effort on the more difficult task. Hence, their performance on the learning outcome (i.e., comprehension test) was unsatisfactory. To put it simply, when the intrinsic load increased, the comprehension was interrupted and the participants invested less effort in accomplishing the tasks.

As to the realm of testing, Pouw et al. (2016) explored the influence of meaningful versus non-meaningful conditions of physical engagement on different forms of competency (i.e., unreflective, reflective, and motoric) in solving problems. No statistically significant correlations between response time and measures of mental effort and perceived difficulty were reported.

Contrasting with the substantial body of evidence on the role of cognitive load in language teaching and learning, research on item loads has received scant attention (Ponce et al., 2020). In general, test items have attracted the researchers' attention in such fields as chemistry (Prisacari & Danielson, 2017), mathematics (Gvozdenko & Chambers, 2007), and algebra (Sweller et al., 2011); however, examining the load of language test items has been restricted to the comparison of item loads between different modes (e.g., computer vs. paper) or types of materials (e.g., static vs. animated graphics) (Dindar et al., 2015; Prisacari & Danielson, 2017). Specifically, the load of multiple-choice language items has been under-explored (Ponce et al., 2020).

Also, there are inconsistencies in the relationship between mental effort and response time. It has been suggested that the longest response time was recorded for the most mental-effort-consuming task (Sasayama, 2016). On the other hand, a moderate correlation (DeLeeuw & Mayer, 2008) and no statistically significant correlation (DeLeeuw & Mayer, 2008; Dindar et al., 2015; Pouw et al., 2016)

between mental effort and response time were reported. Recall that a significant correlation between response time and mental effort was observed in the first experiment of the study conducted by DeLeeuw and Mayer (2008); however, in their second experiment, a non-significant but small correlation was observed. Consequently, there seems to be no consensus among the researchers.

On the other hand, much of the current literature centered on the relationship between task difficulty and mental effort. Many researchers unanimously emphasized a positive relationship between mental effort and task difficulty (Lee, 2019; Révész et al., 2016; Sasayama, 2016). Likewise, DeLeeuw and Mayer (2008) highlighted the significant correlation between mental effort and task complexity. This is in contrast to Lee's (2014) argument.

Hence, due to the scarcity of research on the relationship between difficulty and the cognitive load of language test items, there is a call for the implementation of multiple measures of cognitive load. Specifically, to collect dependable evidence on the individuals' performance, not affected by the load of another task, the use of response time has been suggested. Furthermore, the use of objective and subjective measures can reveal the difference between the actual and the perceived cognitive load.

The paucity of research on the investigation of item functioning as to the measures of cognitive load in the realm of language testing in the Iranian context has incurred several problems because disregarding the cognitive load imposed by language test items on the test taker's minds may lead to imposing a high load on their minds. This can consequently influence their cognitive processes and performance due to the working memory capacity's reaching its limitation (Goldhammer et al., 2014; Ponce et al., 2020; Sweller et al., 1998). Stated another way, success or failure in the test heavily relies on the load of items imposed on the examinees' minds. Test items may not provoke loads commensurate with their designed features and difficulties, which can lead to the overload of test takers' working memory capacity. This can result in the failure in answering an item.

The Present Study

The current study strived to provide a more precise image of load patterns of language test items (i.e., vocabulary and grammar sections) of the MA Iranian

university entrance examination (IUEE) of English majors. Given that item loads can influence test takers' performance and can also reveal information about item functioning, the patterns of item loads were portrayed through the simultaneous implementation of subjective (i.e., perceived mental effort and perceived level of difficulty self-reports) and objective behavioral (i.e., response time) measures.

Stated more precisely, the present study investigated the relationship between response time and mental effort to check if response time can reveal cognitive load in line with mental effort. Also, the relationships between the subjective measures of perceived difficulty and mental effort were sought to check whether both measures match and assess difficulty similarly. Further, the relationships between response time and perceived level of difficulty were explored to ascertain whether the experienced difficulty was reflected in the time spent on answering the test items.

The scrutiny of language test items has blossomed in the world; however, the load of language test items of the MA English majors of the Iranian university entrance examination (IUEE) has not been explored in our context. Therefore, there is a need for methods triangulation (Ary et al., 2019) through subjective and behavioral measures to cast light on the actual cognitive processing and the test item functioning.

The present study strived to address the following questions:

1. Is there any statistically significant relationship between the test-takers' perceived mental effort and the response time for each language test item?
2. Is there any statistically significant relationship between the test takers' perceived mental effort and perceived level of difficulty?
3. Is there any statistically significant relationship between the test takers' perceived level of difficulty and the response time for each language test item?

Method

Participants

Twenty-five male and 35 female MA graduates and students of the University of Tehran, Allameh Tabatabaee University, and Alzahra University, aged

between 21 and 39 (M = 27.28, SD = 4.41) and majoring in teaching English as a foreign language, translation studies, and English literature, attended the study. The participants were selected through convenience sampling (Dornyei, 2007). They were homogenous in terms of language proficiency such that their proficiency level fell into the categories of advanced and very advanced users of the English language after taking the Oxford placement test (OPT) (Dave, 2004). In addition, the piloting phase was necessary to set the fixed timing for presenting each language test item on the screen. To this end, five female and male participants attended the pilot phase. Moreover, regarding the probable influence of practice effect, as the time span between the entrance examination and the time the test was run was approximately two years, it was supposed that this effect would be very unlikely to exist.

Instruments

Multiple-choice vocabulary and grammar items, two self-report ratings, and a proficiency test were the instruments used to collect data. The current study made use of grammar (20 items) and vocabulary (40 items) sections of the MA English majors of the IUEE tests held in 2018 and 2019. The Psychopy software collected the response answer and response time of every multiple-choice question with the precision of milliseconds (Peirce et al., 2019). The present study included two subjective self-reports of perceived mental effort and perceived level of difficulty. As one of the load components, mental effort, developed and validated by Paas (1992), was measured subjectively through the *Mental Effort* rating scale. Examinees can report mental effort on a 7-point symmetrical category scale on a numerical value spanning from *very low* (1) to *very high* (7) mental effort. The reasoning behind the prevalent implementation of mental effort self-report is the simplicity of data gathering and analysis (Paas, 1992; Paas & van Merriënboer, 1993). Further, the reliability of mental effort measure has shown to be acceptable (α = .82).

As for the second rating scale, *Level of Difficulty*, developed and validated by Bratfisch et al. (1972), was rated by a 7-point scale spanning from 1 (*very easy*) to 7 (*very difficult*). Perceived difficulty is concerned with the difficulty of the item itself. This self-report rating scale has also been shown to be a sound indicator of cognitive load (Prisacari & Danielson, 2017). Although task difficulty and mental

effort may have a relationship with each other, they measure different constructs: Task difficulty corresponds to the task itself, and mental effort relates to a process involving more aspects than being limited to the task itself (van Gog & Paas, 2008). To check the homogeneity of participants with respect to language proficiency, the grammar section of the OPT was administered. Based on whether their scores fell in the range of 75 to 100, the categories of advanced to very advanced language users, the participants attended the main phase of the study.

Procedure

Participants took the test for about one hour. They were instructed how to deal with the self-ratings, and had received complete explanations on the definitions of mental effort and perceived difficulty level. They were also directed to answer the language test items as fast and as correctly as possible. Upon confirmation of their understanding of the explanations, the participants were then required to answer the multiple choice test items in addition to the self-reports of mental effort and level of difficulty. Indeed, they answered grammar and vocabulary items sequentially presented in two conditions, each including 20 and 40 items, respectively. The order of language test items was randomized by the Psychopy software. A fixation cross (i.e., 500 ms) was also shown after the presentation of each language test item. Having answered each language item, they rated their perceived difficulty and the amount of experienced mental effort.

Results

Before running Pearson and Spearman rho correlations, the normality assumption for response time, perceived mental effort, and task difficulty was evaluated by Shapiro–Wilk's test ($p > 0.05$). Indeed, no violation of this assumption was observed. The data were also checked for outliers with respect to response time, mental effort, and perceived difficulty. Only two cases' response times fell three standard deviations away from the mean and the overall mean replaced their response times. The reason for running Spearman rho correlation was the fact that the data were collected through Likert-type items. Indeed, Likert-type items or ranked data can sometimes be regarded as ordinal in nature (Boone et al., 2014; Pallant, 2016).

The first research question scrutinized the relationship between the test

takers' perceived mental effort and response time. To answer the question, Pearson correlation and Spearman rho correlation were obtained. However, due to the similarity of the results, only the results pertinent to Pearson correlation are reported. Note that only correct responses are considered in the analysis in the case of response time because only correct answers seem to reveal cognitive load (Lee, 2019). The results from the correlation between response time and mental effort are summarized in Tables 1 and 2. Also note that the results of each grammar and vocabulary section are presented in two separate sets such that the first set represents the items related to the test held in 2019 and the second set indicates the items associated with the test administered in 2018. This can potentially help cross-validate the results through a second data set.

The results obtained from the correlational analyses are set out in Table 1. As evident, a strong positive correlation between the response time and mental effort of the first set of grammar items was observed ($r = .625$, $p < .001$). Additionally, the correlation between response time and mental effort of the second set of grammar items was also significant ($r = .631$, $p < .001$).

Table 1

Pearson Product Moment Correlation of Response Time and Mental Effort (Grammar Items)

			1	2	3	4
1.	Response time grammar items (first set)	Pearson	—	.625**	.640**	.202
		Sig. (2-tailed)		.000	.000	.121
2.	Mental effort grammar items (first set)	Pearson		—	.389**	.566**
		Sig. (2-tailed)			.002	.000
3.	Response time grammar items (second set)	Pearson			—	.631**
		Sig. (2-tailed)				.000
4.	Mental effort grammar items (second set)	Pearson				—
		Sig. (2-tailed)				

Table 2 provides the correlations between mental effort and response time of 40 vocabulary items. This table is quite revealing in two ways. The results indicate that response time and mental effort of the first set of vocabulary items were positively correlated (r = .704, p < .001). A strong positive correlation was also found between mental effort and response time of the second set of vocabulary items (r = .769, p < .001).

Table 2

Pearson Product Moment Correlation of Response Time and Mental Effort (Vocabulary Items)

		1	2	3	4
1. Mental effort vocabulary items (first set)	Pearson	__	.704**	.687**	.515**
	Sig. (2-tailed)		.000	.000	.000
2. Response time vocabulary items (first set)	Pearson		__	.527**	.780**
	Sig. (2-tailed)			.000	.000
3. Mental effort vocabulary items (second set)	Pearson			—	.769**
	Sig. (2-tailed)				.000
4. Response time vocabulary items (second set)	Pearson				—
	Sig. (2-tailed)				

The next research question delved into the relationship between perceived mental effort and level of difficulty of the language test items. To this end, Spearman rho correlation was run as both of the variables were ordinal in nature. Note that the data were analyzed in two ways: considering only correct responses in one case and all responses (i.e., both correct and incorrect answers) in the other case. This was because no evidence was found in the literature for excluding incorrect responses from data analyses. Although results of the correlational analyses of only correct and all responses (i.e., including both incorrect and correct answers) were not too much different, their tables are reported.

As shown in Table 3, a strong positive correlation between mental effort and level of difficulty of the first set of grammar items was detected (r = .891, p < .001) considering only correct responses. It can further be seen from the data

presented in the table that level of difficulty was significantly correlated with perceived mental effort of the second set of grammar items (r = .907, $p < .001$). As evident in Table 4, a strong positive correlation was found between the aforementioned variables in the case of the first set of grammar items including both correct and incorrect responses (r = .839, $p < .001$). Also, looking at Table 4, it is obvious that a strong positive correlation was observed between the aforesaid variables of the second set of grammar items considering all responses (r = .771, $p < .001$). Hence, it appears that the magnitudes of correlations were reduced in the case of including both correct and incorrect answers.

Table 3

Spearman Rho Correlation of Mental Effort and Difficulty Level (Only Correct Answers to Grammar Items)

		1	2	3	4
1. Mental effort grammar items (first set)	Spearman's rho	__	.891**	.556**	.517**
	Sig. (2-tailed)		.000	.000	.000
2. Difficulty level grammar items (first set)	Spearman's rho		__	.523**	.588**
	Sig. (2-tailed)			.000	.000
3. Mental effort grammar items (second set)	Spearman's rho			__	.907**
	Sig. (2-tailed)				.000
4. Difficulty level grammar items (second set)	Spearman's rho				—
	Sig. (2-tailed)				

Table 4

Spearman Rho Correlation of Mental Effort and Difficulty Level (Correct and Incorrect Answers to Grammar Items)

			1	2	3	4
1.	Mental effort grammar items (first set)	Spearman's rho Sig. (2-tailed)	—	.839**	.881**	.743**
				.000	.000	.000
2.	Difficulty level grammar items (first set)	Spearman's rho Sig. (2-tailed)		—	.718**	.892**
					.000	.000
3.	Mental effort grammar items (second set)	Spearman's rho			—	.771**
		Sig. (2-tailed)				.000
4.	Difficulty level grammar items (second set)	Spearman's rho Sig. (2-tailed)				—

Table 5 presents the correlations between perceived mental effort and level of difficulty of vocabulary section considering only correct responses. Looking at the table below, a positive correlation was found between the aforementioned variables of the first set of vocabulary items (r = .792, $p < .001$). Besides, perceived mental effort was significantly correlated with perceived difficulty level of the second set of vocabulary items (r = .881, $p < .001$). Furthermore, Table 6 displays an overview of the aforementioned variables' correlations. Looking at Table 6, it is apparent that a strong positive correlation was observed between the variables of the first set of vocabulary items ($r = .689$, $p < .001$). The results, as evident in Table 6, reveal that a significant positive correlation between the variables of the second set of vocabulary items was found ($r = .724$, $p < .001$). Thus, it can also be concluded that the consideration of all responses can reduce the magnitudes of correlations to some extent.

Table 5

Spearman Rho Correlation of Mental Effort and Difficulty Level (Only Correct Answers to Vocabulary Items)

		1	2	3	4
1. Mental effort vocabulary items (first set)	Spearman's rho	__	.792**	.617**	.493**
	Sig. (2-tailed)		.000	.000	.000
2. Difficulty level vocabulary items (first set)	Spearman's rho		__	.542**	.681**
	Sig. (2-tailed)			.000	.000
3. Mental effort vocabulary items (second set)	Spearman's rho			__	.881**
	Sig. (2-tailed)				.000
4. Difficulty level vocabulary items (second set)	Spearman's rho				
	Sig. (2-tailed)				—

Table 6

Spearman Rho Correlation of Mental Effort and Difficulty Level (Correct and Incorrect Answers to Vocabulary Items)

		1	2	3	4
1. Mental effort vocabulary items (first set)	Spearman's rho	__	.689**	.953**	.727**
	Sig. (2-tailed)		.000	.000	.000
2. Difficulty level vocabulary items (first set)	Spearman's rho		—	.624**	.918**
	Sig. (2-tailed)			.000	.000
3. Mental effort vocabulary items (second set)	Spearman's rho			—	.724**
	Sig. (2-tailed)				.000
4. Difficulty level vocabulary items (second set)	Spearman's rho				
	Sig. (2-tailed)				—

The third research question explored the relationship between the test takers' perceived difficulty level and response time. To answer the proposed

question, it was necessary to obtain both Pearson and Spearman rho correlations because perceived level of difficulty may also be viewed as an ordinal variable. As the results were akin, only those pertinent to Pearson correlation are reported. Recall that incorrect responses are discarded.

The outcomes of the correlational analyses are presented in Tables 7 and 8. As shown in Table 7, there was a strong positive correlation between perceived level of difficulty and response time of the first set of grammar items (r = .578, p < .001). Besides, looking at Table 7, it is evident that there was a significant positive correlation between the aforesaid variables of the second set of grammar items (r = .602, p < .001).

Table 7

Pearson Product Moment Correlation of Response Time and Difficulty Level (Grammar Items)

		1	2	3	4
1. Response time grammar items (first set)	Pearson	—	.578**	.640**	.236
	Sig. (2-tailed)		.000	.000	.069
2. Difficulty level grammar items (first set)	Pearson		—	.338**	.583**
	Sig. (2-tailed)			.008	.000
3. Response time grammar items (second set)	Pearson			—	.602**
	Sig. (2-tailed)				.000
4. Difficulty level grammar items (second set)	Pearson				—
	Sig. (2-tailed)				

Table 8 displays the correlation between perceived level of difficulty and response time for vocabulary items. As this table shows, a positive correlation was found between the mentioned variables of the first set of vocabulary items (r =. 651, p < .001). Also, a strong positive correlation between the variables of the second set of vocabulary items was observed (r =. 785, p < .001).

Table 8

Pearson Product Moment Correlation of Response Time and Difficulty Level (Vocabulary Items)

			1	2	3	4
1.	Response time vocabulary items (first set)	Pearson	__	.651**	.780**	.474**
		Sig. (2-tailed)		.000	.000	.000
2.	Difficulty level vocabulary items (first set)	Pearson		__	.570**	.737**
		Sig. (2-tailed)			.000	.000
3.	Response time vocabulary items (second set)	Pearson			__	.785**
		Sig. (2-tailed)				.000
4.	Difficulty level vocabulary items (second set)	Pearson				
		Sig. (2- tailed)				—

Discussion

The first research question sought to examine the relationship between response time and mental effort. By running Pearson and Spearman rho correlations, statistically significant correlations were observed. The outcomes seem to be contrary to those of DeLeeuw and Mayer (2008), Dindar et al. (2015), and Pouw et al. (2016) who found no significant correlation between response time and mental effort. Note that the results obtained in DeLeeuw and Mayer's (2008) research revealed a correlation with a small effect size of .12. Nonetheless, in part of their study, they found a statistically significant correlation between response time and mental effort. Hence, the current study's findings seem to be partly in accord with DeLeeuw and Mayer's (2008) study outcomes. The reason behind the conflicting findings may lie in making use of instruments which were different in nature.

Our findings also mirror those of Sasayama's (2016) study in which the least and most demanding tasks required the least and most amount of time, respectively. Besides, the easiest and the most difficult tasks were perceived as the least and most mental effort consuming, respectively. Our findings also seem to be in agreement with the outcomes of the recent study carried out by Ponce et al. (2020) who

regarded response time as an appropriate indicator of cognitive load. They further concluded that the higher the cognitive load, the larger the response time needed to answer the question. Hence, response time and mental effort seem to be the sound indicators of cognitive load. Response time can indeed provide some evidence as to how deep the processing is or how much cognitive resources are required to accomplish a task (Goldhammer et al., 2014).

The second research question investigated the relationship between perceived mental effort and task difficulty through running Spearman rho correlation. Significant correlations with large effect sizes were observed with regard to both grammar and vocabulary items. The outcomes of the current study corroborate the findings of numerous studies (e.g., Lee, 2019; Révész et al., 2016; Sasayama, 2016) in which more difficult tasks were perceived as more demanding and required more mental effort. However, our findings are in disagreement with those of the study conducted by Lee (2014) in which the researcher reported when the task at hand became too demanding (i.e., when perceived to have a high level of difficulty), individuals ceased to invest mental effort in accomplishing the task. That is, when encountered with a very difficult task, they became reluctant to make an attempt to complete the task.

The last question explored the relationship between perceived level of difficulty and response time. Statistically significant correlations were detected between the variables of both grammar and vocabulary items. The outcomes of the current study appear to be consistent with those of Sasayama (2016) and Lee (2019) who reported that the most difficult task required the longest amount of time. Our outcomes are not in line with those of the studies carried out by Pouw et al. (2016), Révész et al. (2016), and Révész et al. (2014) who found no significant correlation between perceived level of difficulty and response time.

Directions for Further Research

This study cast light on the role of cognitive load in exploring the language test items to unravel the cognitive processes underlying test taking. However, this study has some limitations that can open up the line for further research. One limitation of this study concerns the mere focus on vocabulary and grammar items. Hence, the cloze test and reading comprehension sections of the entrance

examination were not taken into account due to the execution constraints of the Psychopy software. Another limitation lies in the small number of participants due to the pandemic situation. Several caveats need to be noted about the generalizability of findings concerning the participants. First, the current study has considered only MA students majoring in teaching English as a foreign language, literature, and translation studies with a high level of proficiency. Note that test takers with a high level of proficiency can better distinguish nuances of task difficulty compared to those with a low level of proficiency (Ayres, 2006; Sasayama, 2016). Future studies should include a low proficiency group as well and compare the performance and perceptions of participants with those with a high level of proficiency.

Also, to gain a more profound understanding of the specific criteria that the test takers used to rate the difficulty of each item, retrospective interviews and think-aloud techniques are strongly recommended. Classification of items into groups of the least- to the most- complex ones and comparing their differences can also lead to interesting findings in future investigations. The inequivalent number of characters of grammar items can also be considered a confounding variable. Hence, future studies should address this issue by including grammar items of approximately equal number of characters.

Moreover, the present study focused on measuring mental effort merely through self-report rating scales. To develop a deeper understanding of the influence of task difficulty on mental effort and response time as well as perceived difficulty, a variety of nonintrusive physiological measures such as electroencephalography (Antonenko & Niederhauser, 2010) and eye-tracking (Scheiter et al., 2020) can be applied to capture cognitive load while test takers are taking the test. Ultimately, as test takers themselves might not rate their perceived mental effort or perceived level of difficulty based on consistent reasoning, further studies should be carried out to investigate expert judgments as well (Révész et al., 2016).

Conclusion

In conclusion, the results of the present study indicated that the expected relationships and hypotheses were borne out. That is, the outcomes of the statistical analyses provide support for the predictions that objective and subjective measures

of cognitive load can reveal that more difficult language items impose a greater amount of cognitive load. Hence, both subjective and objective measures seem to match with respect to the difficulty of language items. Also, subjective measures assess cognitive load in a similar way.

The current study can be of paramount significance from various perspectives. Cognitive load theory can contribute to the exploration of item functioning in psychometrics through the concurrent use of various cognitive load measures. In this way, test designers can have a thorough grasp of load and function of the items they develop. This awareness might urge test designers to proceed with caution in designing test items when considering the possible detrimental effects of item malfunctioning on the test takers' minds. In other words, they can examine whether the test items they design correspond to the characteristics being experienced or perceived by the test takers; for example, when it comes to the difficulty of items, the speed of processing, and the investment of mental effort. To this end, the cognitive load measures can provide worthwhile evidence. In this respect, the examination of language test items of MA English majors of the Iranian university entrance examination (IUEE) can provide insights into the item functioning. Ignoring cognitive load measures can make our understanding of item functioning inadequate.

On the other hand, due to some criticism leveled against the use of subjective measures such as being subject to under- or over-estimation of individuals, the behavioral measures can contribute to CLT through providing helpful information about item difficulty, cognitive processes, and item functioning (de Jong, 2010; Ponce et al., 2020; van der Linden, 2009). All in all, the investigation of test item functioning through the implementation of cognitive load measures can be considered an initial stage in the scrutiny of language test items from the cognitive load perspective.

References

Antonenko, P. D., & Niederhauser, D. S. (2010). The influence of leads on cognitive load and learning in a hypertext environment. *Computers in Human Behavior*, *26*(2), 140-150. https://doi.org/10.1016/j.chb.2009.10.014

Ary, D., Jacobs, L. C., Irvine, S., & Walker, D. (2019). *Introduction to research in education* (10th ed.). Wadsworth Cengage Learning.

Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction, 16*(5), 389-400. https://doi.org/10.1016/j.learninstruc.2006.09.001

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer.

Bratfisch, O., Borg, G., & Dornic, S. (1972). *Perceived item-difficulty in three tests of intellectual performance capacity* (Report No. 29). Institute of Applied Psychology.

Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*(1)*,* 53-61. https://doi.org/10.1207/S15326985EP3801_7

Brünken, R., Seufert, T., & Paas, F. (2010). Measuring cognitive load. In J. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 181-202). Cambridge University Press. https://doi.org/10.1017/CBO9780511844744.011

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293-332. https://doi.org/10.1207/s1532690xci0804_2

Dave, A. (2004). *Oxford placement test*. Oxford University Press.

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science, 38*(2), 105-134. https://doi.org/10.1007/s11251-009-9110-0

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology, 100*(1), 223-234. https://doi.org/10.1037/0022-0663.100.1.223

Dindar, M., Yurdakul, I. K., & Dönmez, F. I. (2015). Measuring cognitive load in test items: Static graphics versus animated graphics. *Journal of Computer Assisted Learning, 31*(2), 148-161. https://doi.org/10.1111/jcal.12086

Dornyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.

Gass, S. M., Behney, J., & Plonsky, L. (2013). *Second language acquisition: An introductory course* (4th ed.). Routledge.

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology*, *106*(3), 608-626. https://doi.org/10.1037/a0034716

Gvozdenko, E., & Chambers, D. (2007). Beyond test accuracy: Benefits of measuring response time in computerised testing. *Australasian Journal of Educational Technology*, *23*(4), 542-558. https://doi.org/10.14742/ajet.1251

Johannsen, G. (1979). Workload and workload measurement. In N. Moray (Ed). *Mental workload: Its theory and measurement,* (pp. 3-11). Springer.

Lee, H. (2014). Measuring cognitive load with electroencephalography and self-report: Focus on the effect of English-medium learning for Korean students. *Educational Psychology, 34*(7), 838-848. https://doi.org/10.1080/01443410.2013.860217

Lee, J. (2019). Task complexity, cognitive load, and L1 speech. *Applied Linguistics*, *40*(3), 506-539. https://doi.org/10.1093/applin/amx054

Leppink, J. (2017). Cognitive load theory: Practical implications and an important challenge. *Journal of Taibah University Medical Sciences, 12*(5), 385-391. https://doi.org/10.1016/j.jtumed.2017.05.003

Martin, S. (2014). Measuring cognitive load and cognition: Metrics for technology enhanced learning. *Educational Research and Evaluation: An International Journal on Theory and Practice*, *20*(8), 592-621. https://doi.org/10.1080/13803611.2014.997140

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97. https://doi.org/10.1037/h0043158

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429-434. https://doi.org/10.1037/0022-0663.84.4.429

Paas, F., Ayres, P., & Pachman, M. (2008). Assessment of cognitive load in multimedia learning environments: Theory, methods, and applications. In D. Robinson & G. Schraw (Eds.), *Recent innovations in educational technology that facilitate student learning* (pp. 11-35). Information Age Publishing.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1)*,* 1-4. https://doi.org/10.1207/S15326985EP3801_1

Paas, F., Tuovinen, J. E., Tabbers, H., & van Gerven, P. W. M. (2003). Cognitive load measurements as a means to advance cognitive load theory. *Educational*

Psychologist, *38* (1), 63-71. https://doi.org/10.1207/S15326985EP3801_8

Paas, F., & van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors, 35*(4), 737-743. https://doi.org/10.1177/001872089303500412

Pallant, J. (2016). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (6th ed.). George Allen & Unwin.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195-203. https://doi.org/10.3758/s13428-018-01193-y

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*(3), 193-198. https://doi.org/10.1037/h0049234

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction, 12*(1), 61-86. https://doi.org/10.1016/S0959-4752(01)00016-0

Ponce, H. R., Mayer, R. E., Sitthiworachart, J., & Lopez, M. J. (2020). Effects on response time and accuracy of technology-enhanced cloze tests: An eye-tracking study. *Educational Technology Research and Development*, *68*, 2033-2053. https://doi.org/10.1007/s11423-020-09740-1

Pouw, W. T., Eielts, C., van Gog, T., Zwaan, R. A., & Paas, F. (2016). Does (non)meaningful sensori-motor engagement promote learning with animated physical systems? *Mind, Brain, and Education*, *10*, 91-104. https://doi.org/10.1111/mbe.12105

Prisacari, A. A., & Danielson, J. (2017). Computer-based versus paper-based testing: Investigating testing mode with cognitive load and scratch paper use. *Computers in Human Behavior, 77,* 1-10. https://doi.org/10.1016/j.chb.2017.07.044

Révész, A., Michel, M., & Gilabert, R. (2016). Measuring cognitive task demands using dual-task methodology, subjective self-ratings, and expert judgments: A validation study. *Studies in Second Language Acquisition, 38*(4), 703-737. https://doi.org/10.1017/S0272263115000339

Révész, A., Sachs, R., & Hama., M. (2014). The effects of task complexity and input frequency on the acquisition of the past counterfactual construction through recasts. *Language Learning, 64*, 615-650. https://doi.org/10.1111/lang.12061

Sasayama, S. (2016). Is a 'complex' task really complex? Validating the assumption of cognitive task complexity. *The Modern Language Journal*, *100*, 231-254. https://doi.org/10.1111/modl.12313

Scheiter, K., Ackerman, R., & Hoogerheide, V. (2020). Looking at mental effort appraisals through a metacognitive lens: Are they biased? *Educational Psychology Review*,

32(4)*,* 1003-1027.

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19*(4), 469-508. https://doi.org/10.1007/s10648-007-9053-4

Shohamy, E., Dinitsa-Schmidt, S,. &Ferman, I. (1996). Test impact revisited washback effect over time. *Language Testing*, *13*(3), 298-317. https://doi.org/10.1177/026553229601300305

Skulmowski, A., & Rey, G. D. (2017). Measuring cognitive load in embodied learning settings. *Frontiers in Psychology*, *8*, 1-6. https://doi.org/10.3389/fpsyg.2017.01191

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295-312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J. (1999). *Instructional design in technical areas*. ACER.

Sweller, J. (2010). *Cognitive load theory: Recent theoretical advances.* In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 29-47). Cambridge University Press. https://doi.org/10.1017/CBO9780511844744.004

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development, 68*(1), 1-16. https://doi.org/10.1007/s11423-019-09701-3

Sweller, J., Ayres, P., Kalyuga, S. (2011). *Cognitive load theory.* Springer.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-296. https://doi.org/10.1023/A:1022193728205

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31,* 261-292.

van der Linden, W. J. (2009). Conceptual issues in response-time modeling. *Journal of Educational Measurement, 46*, 247-272. https://doi.org/10.1111/j.1745-3984.2009.00080.x

van Gerven, P. W. M., Paas, F., van Merriënboer, J. J. G., & Schmidt, H. G. (2006). Modality and variability as factors in training the elderly. *Applied Cognitive Psychology, 20*(3), 311-320. https://doi.org/10.1002/acp.1247

van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist*, *43*(1), 16-26. https://doi.org/10.1080/00461520701756248

Whelan, R. R. (2007). Neuroimaging of cognitive load in instructional multimedia. *Educational Research Review, 2*(1),1-12. https://doi.org/10.1016/j.edurev.2006.11.001

Wiebe, E. N., Roberts, E., & Behrend, T. S. (2010). An examination of two mental workload measurement approaches to understanding multimedia learning. *Computers in Human Behavior*, *26*(3), 474-481. https://doi.org/10.1016/j.chb.2009.12.006

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: Mental workload in ergonomics. *Ergonomics, 58*(1), 1-17. https://doi.org/10.1080/00140139.2014.956151

Journal of Language Horizons

Scientefic Quarterly Journal of Language Horizons, *Alzahra University*
Volume 7, Issue 1, Spring 2023 (Serial No. 15)

The Impact of Self-Regulated Instruction on Reading Comprehension and Willingness to Admit Wrongness in Face-to-Face vs. Online Classes

Research Article
pp. 35-58

Mohammad Reza Zebardast[1]
Mehry Haddad Narafshan[*2]
Hassan Shahabi[3]

Received: 2021/06/29 Accepted: 2022/02/12

Abstract

The emerging technologies and their applications in education necessitate attempts to establish and test theories that can affect the optimal benefit of student learning and teacher training. Self-regulated learning theory is one of the most influential theories that recognize the interconnections of different factors for optimal learning as an ongoing process. Therefore, this study attempted to probe the efficiency of self-regulated learning to develop learners' reading comprehension and willingness to admit wrongness in an online learning context in addition to the face-to-face learning environment. To address this issue, a quasi-experimental design was utilized. The experimental groups were instructed to be self-regulated through receiving self-regulated learning strategies, while the control group received no self-regulated instructions. The statistical analysis of the paired sample t-test indicated that just the experimental groups exhibited significant improvement in being empowered with higher levels of reading comprehension and willingness to admit wrongness. Additionally, as the ANCOVA analysis showed, the online group had a better performance

* Corresponding Author
[1] Ph.D. Candidate, Department of Foreign Languages, Kerman Branch, Islamic Azad University, Kerman, Iran. aseman2st@gmail.com
[2] Assistant Professor, Department of Foreign Languages, Kerman Branch, Islamic Azad University, Kerman, Iran. haddad@iauk.ac.ir
[3] Assistant Professor, Department of Foreign Languages, Kerman Branch, Islamic Azad University, Kerman, Iran. hshahabi@iauk.ac.ir

DOI: 10.22051/lghor.2022.36718.1517
DOR: 20.1001.1.2588350.2023.7.1.2.0

compared with the control and face-to-face group regarding reading comprehension skills. The findings tend to direct the attention of researchers, teachers, and course designers toward the importance of using self-regulation models as a possible way to move away from traditional teaching.

Keywords: face-to-face classrooms, online classrooms, reading comprehension, self-regulation, willingness to admit wrongness

Introduction

A cyclical and dynamic process in which learners take control of their feelings, thoughts, and behaviors to reach their learning goals is defined as self-regulated learning (Schunk & Zimmerman, 2010). Self-regulated learning (SRL) involves effective management, setting goals, planning, and reflecting on the learning process (Wong et al., 2020; Zimmerman, 2011). In conventional learning environments, the most effective learners are those who can effectively self-regulate (Boekaerts, 1999). Thus, self-regulation has been identified by many researchers as a sign of academic success. In other words, self-regulation has always been one of the key issues in academic settings. Meta-analyses have advocated that, across different educational levels, the relationships between self-regulation and learners' attainment in academic contexts and self-regulation and course outcomes are significantly positive (Boer et al., 2012; Broadbent & Poon, 2015; Dignath & Büttner, 2008; Sitzmann & Ely, 2011). In educational contexts, self-regulated learners prove themselves as agents who can control their learning processes (Bandura, 2001). They are eager when it comes to learning, and they believe that, by employing effective strategies, they will succeed in reaching their educational objectives. Furthermore, self-regulated learners can define their learning goals and can pursue actions of knowledge construction on their own (Seker, 2016). In contrast, learners who are not equipped with SR become blind receivers and imitators of the presented content. As Paul and Binker (1990) pointed out, such learners, are narrow-minded whose thinking is often ambiguous, insignificant, incorrect, irrelevant, and biased.

In the same manner, SRL comes across as a vital capacity in online learning environments that demands learners to have a higher level of autonomy and low levels of direction and facilitation from the teacher (Lehmann et al., 2014).

Unfortunately, most learners have difficulties employing SRL skills, when confronted with challenging and compound tasks in online learning settings (Azevedo, 2005; Wolters & Hussain, 2015). Cho and Heron (2015) mentioned that, in online environments, students are unmotivated and lack sufficient thinking skills, cognitive strategies, and self-monitoring ability. Learners with SRL deficiency may misinterpret the autonomy of the online environment and, thereby, may fail in learning skills that are expected in online contexts (Barnard et al., 2009). SRL strategies are effective in online courses giving an upper level of autonomy to students due to the teacher's physical absence. Different studies have shown the fruitfulness of SRL processes in enabling students to be prosperous learners in online environments (Winters et al., 2008). This is also supported by a meta-analysis done by Broadbent and Poon (2015) suggesting the existence of a significantly desirable link between SRL strategies and academic success in online settings. Accordingly, when students are exposed to complex topics that are not accompanied by SRL support, they do not acquire sufficient understanding because they are not adept at regulating their learning (Azevedo & Hadwin, 2005).

Reading is one of the most influential skills in education, and presumably, the most salient one for second language (L2) learners because it can not only boost language learning but also improve other related learning subjects (Nunan, 2003). It is suggested that, if learners are not well-equipped with adequate competency to grasp the tasks, they will probably fail in completing complex tasks (Gerla, 2009; Taraban & Rynearson, 2004; Voge, 2011; Willingham & Price, 2014). Many of them do not own the adequate reading skills that are necessary for them when they are confronted with the rigors of content-area teaching (Beaufort, 2009; Collins et al., 2008; Simsek & Balaban, 2010; Voge, 2011). In this domain, there have been some studies considering the impact of self-regulation on reading attainment in L2 contexts. For instance, Maftoon and Tasnimi (2014) researched this issue, and they found that the execution of self-regulation strategies improved EFL learners' reading comprehension. Other studies also indicated that training learners to be self-regulated on academic content can empower them to succeed in learning (e.g., Kang, 2010; Orhan, 2007).

Additionally, current neo-Vygotskian sociocultural theories on teaching and learning (e.g., Mercer, 2000) suggest that learners' active interaction in the

classroom, both between teachers and students and among students working in pairs, is vital for learning (e.g., Van Lier, 2008). Thus, knowing the structure and functions of this interaction can provide us with a better understanding of how to implement the processes of teaching, and learning in classroom contexts (e.g., Nikula et al., 2013). In this regard, willingness to admit wrongness as a public acceptance that one has been wrong about a belief or attitude was studied in this research. willingness to admit wrongness follows two main features: a changed attitude and public expression (Schumann, 2018).

Overall, assuming reading to be one of the most important aspects of learning (Nunan, 2003), and knowing that self-regulation appears to be critical for academic success (Boekaerts, 1999), the following research questions guide the present study:

1. What is the effect of SRL-based instruction on learners' reading comprehension?
2. What is the effect of SRL-based instruction on learners' willingness to admit wrongness?

Literature Review

Zimmerman's (2000) model contains three cyclic phases; forethought, performance, and self-reflection. The first phase, forethought, refers to processes, such as task analysis and self-motivation beliefs. Task analysis covers goal setting and strategic planning. The former involves assigning specific proximal goals. The latter contains self-efficacy, which refers to learners' beliefs about their learning ability and their outcome expectations of learning. Zimmerman (2000) mentioned two more important terms for this phase to be intrinsic interest, which requires students to value the task skill for its own merits, and learning goal orientation referring to students' belief about the purposes of their learning.

The performance phase is characterized by two procedures of self-control and self-observation that happen during the behavioral implementation. Self-control covers the utilization of the strategic planning of the forethought phase. Other sub-processes of self-control are the use of imagery, self-instruction, attention focusing, and task strategies. As Panadero and Alonso Tapia (2014) referred, the second performance process contains two sub-processes, namely metacognitive monitoring, and self-recording. Meta-cognitive monitoring involves comparing one's

functioning against criteria to evaluate its effectiveness. And self-recording is keeping a record of personal behavior. The third phase, which is called self-reflection, refers to processes that follow each learning effort. It entails two major sub-processes of self-judgment and self-reaction. One form of self-judgment, self-evaluation, refers to the processes in which learners compare their self-observed performances against some criteria, such as their prior performance, another learner's performance, or an absolute standard of performance. Causal attribution is another form of self-judgment that refers to the time when a learner tries to explain one's failure or success. The second process of the self-reflection phase is called self-reaction. Self-reaction embodies the feelings of learners when they reach self-satisfaction or dissatisfaction in connection with their performance. Self-reactions also take the form of adaptive or defensive inferences, which entail decisions about learners' willingness to keep using the same strategies (i.e., adaptive inferences) or avoid performing them (i.e., defensive inferences). This view of self-regulation is cyclical since self-reflection from prior efforts can set the stage for the beginning of a new cycle (Zimmerman, 2000).

Moreover, one of the essential issues in language teaching and learning in today's world is functional mastery of language skills. Many scholars believe reading is the most critical skill in learning English (Bakhshizadeh Gashti, 2018; Janfeshan & Pourarian, 2017; Kharaghani et al., 2016; Mehrpour et al., 2012; Namaziandost et al., 2020). Goldenberg (2011) claims that the primary goal of reading is comprehension, and everything else is a means to this end. Comprehension is the ability to go beyond the words, to understand the ideas in a text and the relationships that exist between those ideas (McNamara, 2007). In this context, reading comprehension is considered one of the essential skills in higher education and the most crucial skill in education (Bedle, 2017; Ghahari & Basanjideh, 2017). Overall, current research suggests that self-regulation correlates with reading comprehension (e.g., Collins et al., 2008). Nevertheless, only a few researchers have provided empirical evidence on the ways self-regulation training can contribute to second/foreign language development and acquisition in general (e.g., Ellis & Zimmerman, 2001; Magno, 2009; Rose & Harbon, 2013) and second/foreign language reading ability in particular (e.g., Finkbeine et al., 2012).

Another important perceptive capacity is Willingness to Admit Wrongness

(WAW). WAW is a public acceptance of a wrong belief or attitude which has changed (Fetterman et al., 2018). In other words, wrongness admission is the act of a person publicly acknowledging that they held an inaccurate belief or attitude. Some people seem more willing to engage in wrongness admission than others. These individual differences may be important in understanding the barriers of wrongness admission. WAW is based on two main features: a changed attitude and the public expression of that attitude (Schumann, 2018). Several studies have reported apologies issues (e.g., Howell et al., 2012; Schumann, 2018), changing attitudes (e.g., see Petty & Briñol, 2015), and being wrong (e.g., see Tavris & Aronson, 2008), but wrongness admission has received little attention among experts and researchers, especially in education-related fields.

Nevertheless, only a few researchers have provided empirical evidence on the ways self-regulation training can contribute to second/foreign language development and also few studies have focused on the role of willingness to admit wrongness. Therefore, understanding the dynamic interaction between these factors, and measuring the enhancing effects of self-regulatory instructions on students' reading comprehension and willingness to admit wrongness demands further studies.

Method

Research Design

This quantitative study followed a quasi-experimental design. The study employed a pre-test and post-test design with two experimental groups and one control group and tried to investigate the effect of experimental treatment, which was self-regulation instructions in face-to-face and online settings. The dependent variables were learners' reading comprehension, and their willingness to admit wrongness. The participants, as shown in Table 1, were assigned to three groups: control, face-to-face, and online groups.

Table 1

Frequency of Sample Distribution

Groups	Frequency	Percent
Control	13	33.3%
Face-to-Face	13	33.3%
Online	13	33.3%
Total	39	100%

Materials and Instruments

Initial approval was obtained from the institute principals based on the institute's assigned ethical guidelines. Then, a consent form was given to each participant in which the agreement between the researcher and the research participants was outlined, including the roles and responsibilities they were taking towards one another throughout the whole research process. A convenience sampling procedure was used to recruit 39 male and female Iranian EFL learners who ranged in age from 14 to 16 and were homogeneous (Elementary level) based on the results of Longman Placement Test. They were all studying English at the Iran Language Institute of Kerman, Iran. The reading materials were selected from *Select Reading* (Second edition), and participants were exposed to some selected reading texts and related questions. In other words, they were not exposed to or practiced any extra reading texts, tasks, or tests either during the pre-reading or post-reading phases except for the SLR ones. Two reading comprehension tests from *Select Reading* (Second edition) were excluded from the procedural phase and were used as pretests and posttest measures to assess the participants' reading comprehension skills. The reading passages were followed by 8 multiple-choice tests. And 20 minutes were given to the participants to complete the reading questions.

For the other variable which was Willingness to Admit Wrongness, MacCann and Roberts (2008) test was used in this study. The questionnaire items of this test covered different situations (e.g., classroom and home), partners (e.g., stranger, friend, and parent), and contents (e.g., class performance, homework, and discussions). Participants answered if they would admit that they were wrong in each case on a Likert scale. For example, *you are arguing with a classmate (e.g., in a classroom discussion). You are both convinced of your correctness. However, you realize that your opinion is against the facts, and your classmate is probably right. Would you publicly admit your wrongness?*

Before the treatment, all participants of the experimental groups and control group were invited to complete the reading comprehension, and willingness to admit wrongness questionnaires within two separate sessions before the actualization of the research project. Then, in the treatment phase, the control group was exposed to the face-to-face instructions of the ILI institute:

- Answering pre-reading questions
- Learning the passage while the teacher paraphrases the text for them
- Assignment (reading out the passage or a part of it)
- Answering the comprehension questions of the text

Moreover, the learners of the face-to-face group and the online group dealt with the following procedures:

- Answering pre-reading questions
- Completing their forethought checklist.
- Exposing them to the reading content.
- Learning the passage by themselves while completing their performance checklist.
- Answering the comprehension questions of the text
- Completing their reflection checklist

The checklist used in this self-regulatory course was provided based on Morshedian et al.'s (2016) implication of SRL procedures. The following table demonstrates the checklist of the strategies taught and given to the learners in the forethought phase.

Table 2

Checklist of Forethought Phase

Forethought Phase		
Strategy	Forethought Question Prompts	Answer
1. Goal-setting	What is my reading purpose?	
2. Strategic planning	Which strategies should I use during a reading text?	
3. Self-efficacy	How sure am I to get 85 on my next reading test and that I can answer 70% of these reading questions?	
4. Outcome expectations	Do I believe I will cope with the reading task without help?	
5. Task interest/value	How interesting is reading for me?	
6. Goal-orientation	Do I know what the reading goals are?	

The following table demonstrates the checklist of the strategies taught and given to the learners in the performance phase.

Table 3

Checklist of Performance Phase

Performance Phase				
Strategy	Performance Question Prompts	Answer		
1. Modeling	Which reading comprehension strategies am I using?	Scanning	Skimming	Both
2. Self-instruction	Am I saying out loud the strategy/strategies that I am using?			
3. Imagery	Am I using concept mapping?			
4. Time management	Am I following my schedule?			
5. Environmental structuring	Am I in a non-disturbing environment?			
6. Help-seeking	Am I getting help from a capable peer?			
7. Self-consequences	Am I rewarding myself for a better completion			
8. Meta-cognitive monitoring	Are my listed learning features different from the ones mentioned by others (the teachers and classmates)?			
9. Self-recoding	Am I recording the time of the performance?			

And, the following table demonstrates the checklist of the strategies taught and given to the learners in the reflection phase.

Table 4

Checklist of Reflection Phase

Reflection Phase		
Strategy	Performance Question Prompts	Answer
1. Self-evaluation	How well did I understand the reading questions?	
2. Causal attributions	What did cause my poor performance?	
3. Self-reaction	How satisfied am I with my accomplishment on my last reading test/task?	
4. Adaptive/defensive inferences	What do I need to do to boost my performance on my next reading task?	

Therefore, in every reading session, the checklists mentioned above were practiced by learners in the experimental group of the face-to-face and the online group. Moreover, in the experimental group of online, self-regulation instructions were applied with the Adobe Connect program to establish the online teaching context. In the online context, the teacher and learners used microphones and webcam sharing instead of face-to-face communication. Also, these checklists were printed and given to the learners before the intervention program. Table 5 provides a brief description of the self-regulation training sessions in the face-to-face and online contexts.

Table 5

Brief Description of Self-Regulation Training Sessions

Session	Intervention
First to Fifth	Teaching and practicing the forethought phase along with the relevant self-regulatory instructions
Sixth to Tenth	Modeling and practicing the performance phase along with the relevant self-regulatory instructions
Eleventh to Fifteenth	Training and practicing the self-reflection phase along with the relevant self-regulation instructions
Sixteenth to Twentieth	Review and practice all three stages of forethought, execution, and self-reflection

After the treatment, they completed reading comprehension, and willingness to admit wrongness questionnaires. The instructor ensured adherence to the intervention protocol using the following strategies: structured training, supervisory monitoring, and feedback.

Results

The first question of this study aimed at studying whether applying SRL strategies would make a significant difference in learners' reading comprehension or not. To this end, descriptive and inferential statistics such as paired sample t-tests and one-way analysis of covariance were used; Excel 2010 and SPSS 16 software were used, respectively. Table 6 provided descriptive statistics regarding the reading comprehension variable.

Table 6

Descriptive Statistics of the Reading Comprehension Variable in the Pretest and Posttest Phases

Groups	Variable	Time	N	Mean	Std. Deviation
Control	Comprehension	Pre	13	12.12	2
		Post	13	11.69	3.04
Face-to-Face	Comprehension	Pre	13	10.96	2.61
		Post	13	11.38	2.75
Online	Comprehension	Pre	13	10.38	2.67
		Post	13	14.46	3.07

As the descriptive statistics (Table 6) revealed, some differences existed among the means of the pretests and posttests. Therefore, paired sample t-test was used to have a better understanding of these differences (Table 7). As the results of the paired-sample t-test indicated, the control group's reading comprehension scores in the pretest and the posttest did not positively change from the pretest to the posttest. However, the means of reading comprehension scores in the face-to-face and online groups were different in the pretests and the posttests (Table 7).

Table 7

Paired Sample T-Test of Reading Comprehension

Groups	Time	N	Mean	Std. Deviation	T-Test	Df	P-Value
Control	Pre	13	12. 12	2	0.67	12	0.5
	Post	13	11.69	3.04			
Face-to-Face	Pre	13	10.96	2.61	-0.86	12	0.4
	Post	13	11.38	2.75			
Online	Pre	13	10.38	2.67	-4.25	12	0.001
	Post	13	14.46	3.07			

Then the test of between-subject effects was applied and, as the results displayed in Table 8 show, there were no statistically meaningful differences among the three groups on the pre-test reading comprehension scores.

Table 8

Test of Between-Subjects Effects

Source	Sum of Squares	df	Mean Square	F	P-Value
Group	41.61	2	20.81	3.41	0.05
Pretest	104.31	1	104.31	17.11	0.000
Pretest× Group	17.81	2	8.90	1.46	0.2
Error	201.20	33	6.1	-	-

ANCOVA test was used to check the influence of the treatment on the learners' reading comprehension while removing the effect of their pretest scores. Accordingly, levene's test for equality of variances was used (Table 9), and equal variance was assumed.

Table 9

Levene Test of Equality of Error Variances

Variable	F	df1	df2	P-Value
Reading comprehension	1.29	2	36	0.3

The ANCOVA test results showed that the online group performed significantly better than the control and face-to-face group on reading comprehension (Table 10). In addition, the Bonferroni Post Hoc Test was done to uncover specific differences between these three group means. The Bonferroni Post Hoc Test suggested a very large magnitude of instructional effect for the online group as compared to the control and face-to-face group (Table 11).

Table 10

ANCOVA Analysis of Reading Comprehension

Source	Sum of Squares	df	Mean Square	F	P-Value
Pretest	96.07	1	96.07	15.35	0.000
Group	112.70	2	56.35	9.01	0.001
Error	219.01	35	6.26	-	-
Corrected Total	389.74	38	-	-	-

Table 11

Bonferroni Post Hoc Test of Reading Comprehension

Groups	Estimated Marginal Mean	Mean Difference	P-Value
Control	11.05	-0.46	1
Face-to-Face	11.51		
Control	11.05	-3.93	0.002
Online	14.98		
Face-to-Face	11.51	-3.46	0.004
Online	14.98		

The second question of this study implied studying the influence of SRL strategies on learners' willingness to admit wrongness. To investigate it, the following procedures were applied. Table 12 provides descriptive statistics to compare the differences between the means of the control group and experimental groups regarding the willingness to admit wrongness variable.

Table 12

Descriptive Statistics of WAW Variable

Groups	Variable	Time	N	Mean	Std. Deviation
Control	WAW	Pre	13	31.38	4.89
		Post	13	31.77	3.85
Face-to-Face	WAW	Pre	13	29.23	5.51
		Post	13	31.54	5.33
Online	WAW	Pre	13	29.00	4.95
		Post	13	32.38	5.25

As the descriptive statistics (Table 12) revealed, some differences existed among the means of the pretests and posttests. Therefore, paired sample t-test was used to have a better understanding of these differences (Table 13). The result of the control group's willingness to admit wrongness in the pretest and the posttest indicated that willingness to admit wrongness means in the control group did not change from the pretest to the posttest (Table 13). However, the means of willingness to admit wrongness in the face-to-face and online groups were different in the pretests and the posttests (Table 13).

Table 13

Paired Sample T-Test to Compare Pretest and Posttest of WAW

Groups	Time	N	Mean	Std. Deviation	T-Test	df	P-Value
Control	Pre	13	31.38	4.89	-0.79	12	0.4
	Post	13	31.77	3.85			
Face-to-Face	Pre	13	29.23	5.51	-2.30	12	0.04
	Post	13	31.54	5.33			
Online	Pre	13	29.00	4.95	-3.33	12	0.006
	Post	13	32.38	5.25			

The results displayed in Table 14 provided that there were no statistically significant differences among the three groups on the pretest of the willingness to admit wrongness.

Table 14

Test of Between-Subjects Effects

Source	Sum of Squares	df	Mean Square	F	P-Value
Group	0.28	2	0.14	0.02	0.9
Pretest	543.75	1	543.75	59.59	0.00
Pretest× Group	0.32	2	0.16	0.02	0.9
Error	301.13	33	9.13	-	-

To assess the equality of variances for the willingness to admit wrongness, levene's test was used, and equal variance was assumed.

Table 15

Levene Test of Equality of Error Variances

Variable	F	df1	df2	P-Value
WAW	3.27	2	36	0.05

Therefore, ANCOVA was used to check the influence of the treatment on the learners' willingness to admit wrongness while removing the effect of their pretest scores. The ANCOVA test results showed that there were no statistically

significant differences among the three groups on the posttest of the willingness to admit wrongness (Table 16).

Table 16

ANCOVA Analysis of WAW

Source	Sum of Squares	df	Mean Square	F	P-Value
Pretest	549.17	1	549.17	63.76	0.000
Group	37.24	2	18.62	2.16	0.1
Error	301.45	35	8.61	-	-
Corrected Total	855.59	38	-	-	-

Discussion

This research investigated the influence of SRL instruction on learners' reading comprehension and willingness to admit wrongness in two contexts: face-to-face and online. Self-regulation has always been the elemental conceptualization of how to foster learners' progress and is linked to providing environments that push learners' performance toward learning. There is also a line of research that studies the importance of creating a classroom context to pursue educational goals (Alonso-Tapia & Fernandez, 2008; Meece et al., 2006). Following the same objectives, this study tried to assess the effectiveness of the teachability of the latest version of the SRL model (Zimmerman & Moylan, 2009) and its effects on learners' reading comprehension and willingness to admit wrongness in face-to-face and online environments. Drawing on the statistical analysis of the paired sample t-test, the results indicated that reading comprehension and willingness to admit wrongness of face-to-face and online groups changed significantly from the pretest to the posttest. Additionally, as the statistical analysis of the ANCOVA and Bonferroni test showed, the online group had a better performance compared with the control and face-to-face group regarding reading comprehension attainment. However, the ANCOVA test results showed that there were no statistically significant differences among the three groups on the posttest of the willingness to admit wrongness.

The result of this study supports the theoretical dimensions of Zimmerman and Moylan's (2009) model of SRL and supports the applicability of this model to L2 reading classrooms. Moreover, the study provides evidence in favor of the

practicality of effective training of SRL-based reading instruction, which has not been given sufficient consideration in self-regulation studies, especially in online environments. Our findings show the effective instruction of the SRL-based strategies on reading performance as has been mentioned by different researchers (Oxford, 2017; Teng & Zhang, 2020). In the current study, self-regulation strategies also gave the learners the tendency to acknowledge in front of others that they held an inaccurate attitude or belief.

Some other studies have confirmed our findings that using SRL strategies causes cognitive and personal development in different learning contexts (Zhang et al., 2016; Zhang et al., 2019). The findings also advocate that self-regulatory skills help learners control their thoughts, plan their goals, face potential problems, and complete the assigned tasks (McClelland & Cameron, 2019), which in turn equips them to adapt to the academic challenges, needs, and goals. Concerning the significance of SRL strategies in online environments, some studies (Azevedo, 2005; Barnard et al., 2009; Lehmann et al., 2014) have confirmed our findings that effective SRL strategies are critical in online situations. The results are also in line with Broadbent and Poon's (2015) meta-analysis suggesting the existence of a strong interrelationship between self-regulatory strategies and achievements in online settings. Azevedo and Hadwin (2005) have also reported a positive relationship between successful performance in online environments and the effective implantation of self-regulation strategies. In line with our findings, empirical evidence is provided by different researchers in favor of the practical implementation of self-regulation strategies (e.g., Ferreira & Veiga-Simão, 2012; Kang, 2010; Koehler, 2007; Orhan, 2007).

Conclusion

Nowadays, virtual education is becoming increasingly important in both working and learning contexts (Muñoz Cristóbal et al., 2018), and technological development has constantly expanded its prominent role in educational contexts by providing the possibilities for distance education and contributing to the growth of online/virtual education. Among the issues related to online instructional environments, this research compared self-regulation learning in face-to-face learning and online learning. Online learning is nowadays one of the main interests

of educators, teachers, and language learners because it provides a safe and healthy context to access countless information. Therefore, the quasi-experimental design to study the importance of self-regulation instruction was applied, and the following issues were investigated: (a) the teachability of self-regulation strategies, (b) improvement in learners' reading comprehension, and (c) the potential of online context as a more efficient context for learners to be self-regulated readers.

It can be concluded that the findings of this research advocate the previous findings supporting the positive effect of self-regulated instruction on students' academic performance (Boer et al., 2012; Broadbent & Poon, 2015; Büttner, 2008; Dignath et al., 2008; Sitzmann & Ely, 2011), especially in online environments. Furthermore, the results contribute to the literature on e-learning regarding the quality of instructional teaching tasks and learning processes, not the technology itself. In conclusion, the study opens a new window for teachers, educational policymakers, teacher training centers, and those who care about the quality of education and students' academic achievements.

It is noteworthy to consider the present study's limitations. Firstly, limitations of participants' accessibility (elementary language learners) call for further studies demanding students at other institutes and other proficiency levels. Secondly, future studies can consider investigating the interrelationship between some other variables like SRL constructs, language skills, psychological factors, and performing self-regulation in different contexts or sample sizes.

References

Alonso-Tapia J., & Fernandez B. (2008). Development and initial validation of the classroom motivational climate questionnaire (CMCQ). *Psicothema, 20*(4), 883-889. https://www.redalyc.org/articulo.oa?id=727/72720456

Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educational Psychologist, 40*(4), 199-209. https://doi.org/10.1207/s15326985ep4004_2

Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition–Implications for the design of computer-based scaffolds. *Instructional Science*, *33*(5-6), 367-379. https://doi.org/10.1007/s11251-005-1272-9

Bakhshizadeh Gashti, Y. (2018). The effect of authentic and simplified literary texts on the reading comprehension of Iranian advanced EFL learners. *Iranian Journal of English for Academic Purposes*, *7*(2), 32-44. http://journalscmu.sinaweb.net/article_87254.html

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, *52*, 1-26. https://doi.org/10.1146/annurev.psych.52.1.1

Barnard, L., Lan, W. Y., To, Y. M., Paton, V. O., & Lai, S. L. (2009). Measuring self-regulation in online and blended learning environments. *Internet and Higher Education, 12*, 1-6. https://doi.org/10.1016/j.iheduc.2008.10.005

Beaufort, A. (2009). Preparing adolescents for the literacy demands of the 21st century workplace. In L. Christenbury, R. Bomer, & P. Smagorinsky (Eds.), *The handbook of adolescent literacy research,* (pp.239-255). Guilford Press.

Bedle, S. (2017). *Reading strategies in an EFL context: A mixed methods research synthesis* [Master's thesis, University of Glasgow]. https://www.teachingenglish.org.uk/sites/teacheng/files/steven_bedle_university_of_glasgow_dissertation.pdf

Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research*, *31*(6), 445-457. https://doi.org/10.1016/S0883-0355(99)00014-2

Boer, H., Donker-Bergstra, A. S., Kostons, D. D. N. M., Korpershoek, H., & van der Werf, M. P. (2012). *Effective strategies for self-regulated learning: A meta-analysis*. GION onderzoek/onderwijs. https://research.rug.nl/files/54332660/Effective_Strategies_for_Self_regulated_Learning.pdf

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education, 27,* 1-13.

https://doi.org/10.1016/j.iheduc.2015.04.007

Cho, M.-H., & Heron, M. L. (2015). Self-regulated learning: The role of motivation, emotion, and use of learning strategies in students' learning experiences in a self-paced online mathematics course. *Distance Education, 36*(1), 80-99. https://doi.org/10.1080/01587919.2015.1019963

Collins, K. T., Onwuegbuzie, A. J., & Qun G., J. (2008). Reading ability as a predictor of academic procrastination among African American graduate students. *Reading Psychology*, *29*(6), 493-507. https://doi.org/10.1080/02702710802168568

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, *3*, 231-264. https://doi.org/10.1007/s11409-008-9029-x

Dignath, C., Büttner, G., & Langfeldt, H.-P. (2008). How can primary school students learn self- regulated learning strategies most effectively? *Educational Research Review*, *3*, 101-129. https://doi.org/10.1016/j.edurev.2008.02.003

Ellis, D., & Zimmerman, B. J. (2001). Enhancing self-monitoring during self-regulated learning of speech. In H. J. Hartman (Eds.), *Metacognition in learning and instruction*. Springer. https://doi.org/10.1007/978-94-017-2243-8_10

Ferreira, P. C., & Veiga-Simão, A. M. (2012). Teaching practices that foster self-regulated learning: A case study. *Educational Research eJournal*, *1*(1)*,* 1-16. https://doi.org/10.5838/erej.2012.11.01

Fetterman, A. K., Rutjens, B., Landkammer, K., & Wilkowski, B. W. (2018). On postapocalyptic and doomsday prepping beliefs: A new measure, its correlates, and the motivation to prep. *Personality, 33*(4)*,* 506-525. https://doi.org/10.1002/per.2216

Fetterman, A. K., & Sassenberg, K. (2015). The reputational consequences of failed replications and wrongness admission among scientists. *PloS One, 10*(12), 0143723. https://doi.org/10.1371/journal.pone.0143723

Finkbeiner, C., Knierim, M., Smasal, M., & Ludwig, P. H. (2012). Self-regulated cooperative EFL reading tasks: Students' strategy use and teachers' support. *Language Awareness*, *21*(1-2), 57-83. https://doi.org/10.1080/09658416.2011.639892

Gerla, J. (2009). Case study: Jill, a pre-service teacher becomes a reader. *Southeastern Teacher Education Journal*, *2*(2), 145-152. https://digitalcommons.liberty.edu /cgi/viewcontent.cgi?article=2558&context=doctoral

Ghahari, S., & Basanjideh, M. (2017). Psycho-linguistic model of reading strategies awareness in EFL contexts. *Reading Psychology, 38*(2), 125-153.

https://psycnet.apa.org/record/2016-57076-001

Goldenberg, C. (2011). Reading instruction for English language learners. In M. L. Kamil, P. D. Pearson, E. B. Moje & P. P. Afflerbach (Eds.), *Handbook of reading research* (vol. 4) (pp. 684-710). Routledge.

Howell, A. J., Turowski, J. B., & Buro, K. (2012). Guilt, empathy, and apology. *Personality and Individual Differences, 53*(7), 917-922. https://doi.org/10.1016/j.paid.2012.06.021

Janfeshan, K., & Pourarian, M. (2017). The impact of gender on using reading sub skills in Iranian EFL learners reading comprehension. *Journal of Applied Linguistics and Applied Literature: Dynamics and Advances*, *5*(2), 47-60. https://doi.org/10.22049/jalda.2018.26241.1064

Kang, Y. (2010). *Self-regulatory training for helping students with special needs to learn mathematics* [Unpublished doctoral dissertation]. University of Iowa. https://doi.org/10.17077/etd.yucmgqwa

Kharaghani, N., Ghonsooly, B., & Birjandi, P. (2016). The impact of motivation on reading comprehension ability of Iranian EFL learners receiving reciprocal teaching and cooperative grouping intervention program. *Language and Translation Studies (LTS)*, *48*(4), 99-116. https://doi.org/10.22067/lts.v48i4.44503

Koehler, A. (2007). *Raising awareness of self-efficacy through self-regulated learning strategies for reading in a secondary ESL classroom*. School of Education Student Capstone Theses and Dissertations. 349. https://digitalcommons.hamline.edu/hse_all/349

Lee, L., & Gundersen, E. (2011). *Select Readings (Intermediate)*. Oxford University Press

Lehmann, T., Hähnlein, I., & Ifenthaler, D. (2014). Cognitive, metacognitive and motivational perspectives on reflection in self-regulated online learning. *Computers in Human Behavior*, *32,* 313-323. https://doi.org/10.1016/j.chb.2013.07.051

MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: Theory and data. *Emotion*, *8*(4), 540-551. https://doi.org/10.1037/a0012746

Maftoon, P., & Tasnimi, M. (2014). Using self-regulation to enhance EFL learners' reading comprehension. *Journal of Language Teaching and Research*, *5*, 844–855. https://doi.org/10.4304/jltr

Magno, C. (2009). Self-regulation and approaches to learning in English composition writing. *Online Submission*, *1*, 1-16. https://www.researchgate.net/publication/228175957_Self-Regulation_and_Approaches_to_Learning_in_English_Composition_Writing

McClelland, M. M., & Cameron. C. E. (2019). Developing together: The role of executive

function and motor skills in children's early academic lives. *Early Childhood Research Quarterly, 46*, 142-151. https://doi.org/10.1016/j.ecresq.2018.03.014

McNamara, D. S. (2007). *Reading comprehension strategies: Theories, interventions and technologies*. Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203810033

Meece J. L., Anderman E. M., & Anderman L. H. (2006). Classroom goal structure, student motivation, and academic achievement. *Annual Review of Psychology, 57*, 487-503. https://doi.org/10.1146/annurev.psych.56.091103.070258

Mehrpour, S., Sadighi, F., & Bagheri, Z. (2012). Teaching reading comprehension strategies to Iranian EFL pre-university students. *Teaching English as a Second Language Quarterly*, *31*(1), 107-139. https://doi.org/10.22099/jtls.2012.331

Mercer, N. (2000). *Words and minds: How we use language to think together*. Routledge. https://doi.org/10.4324/9780203464984

Morshedian, M., Hemmati, F., Sotoudehnama, E., & Soleimani, H. (2016). The impact of training EFL learners in self-regulation of reading on their EFL literal and critical reading comprehension: Implementing a model. *Journal of Teaching Language Skills*, *35*(2), 99-122. https://doi.org/10.22099/jtls.2016.3788

Muñoz Cristóbal, J. A., Rodríguez-Triana, M. J., Gallego-Lema, V., Arribas-Cubero, H. F., Asensio-Pérez, J. I., & Martínez-Monés, A. (2018). Monitoring for awareness and reflection in ubiquitous learning environments. *International Journal of Human–Computer Interaction, 34*(2), 146-165. https://doi.org/ 10.1080/ 1044 73 18.2017.1331536

Namaziandost, E., Gilakjani, A. P., & Hidayatullah. (2020). Enhancing pre-intermediate EFL learners' reading comprehension through the use of Jigsaw technique. *Cogent Arts & Humanities*, *7*(1), 1738833. https://doi.org/10.1080/23311983.2020.1738833

Nikula, T., Dalton-Puffer, C., & Llinares, A. (2013). CLIL classroom discourse. *Journal of Immersion and Content-Based Language Education*, *1*(1), 70-100. https://doi.org/10.1075/jicb.1.1.04nik

Nunan, D. (2003). *Practical English language teaching*. McGraw-Hill.

Orhan, F. (2007). Applying self-regulated learning strategies in a blended learning instruction. *World Applied Sciences Journal*, *2*(4), 390-398. https://www.academia.edu/260770/Applying_Self_Regulated_Learning_Strategies_In_a_Blended_Learning_Instruction

Oxford, R. L. (2017). *Teaching and researching language learning strategies: Self-regulation in context* (2nd ed.). Routledge. https://doi.org/10.4324/9781315719146

Panadero, E., & Alonso-Tapia, J. (2014). How do students self-regulate? Review of

Zimmerman's cyclical model of self-regulated learning. *Anales de Psicología, 30*(2), 450-462. https://doi.org/10.6018/analesps.30.2.167221

Paul, R., & Binker, A. J. A. (1990). *Critical thinking: What every person needs to survive in a rapidly changing world*. Sonoma State University. https://eric.ed.gov/?id=ED338557

Petty, R. E., & Briñol, P. (2015). Emotion and persuasion: Cognitive and meta-cognitive processes impact attitudes. *Cognition and Emotion, 29*, 1-26. https://doi.org/10.1080/02699931.2014.967183

Rose, H., & Harbon, L. (2013). Self-regulation in second language learning: An investigation of the kanji-learning task. *Foreign Language Annals*, *46*(1), 96-107. https://doi.org/10.1111/flan.12011

Schumann, K. (2018). The psychology of offering an apology: Understanding the barriers to apologizing and how to overcome them. *Current Directions in Psychological Science, 27*(2), 74-78. https://doi.org/10.1177/0963721417741709

Seker, M. (2016). The use of self-regulation strategies by foreign language learners and its role in language achievement. *Language Teaching Research*, *20*(5), 600-618. https://doi.org/10.1177/1362168815578550

Simsek, A., & Balaban, J. (2010). Learning strategies of successful and unsuccessful university student. *Contemporary Educational Technology, 1*(1), 36-45. https://doi.org/10.30935/cedtech/5960

Sitzmann, T., & Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go. *Psychological Bulletin, 137*(3), 421-42. https://doi.org/10.1037/a0022777

Tavris, C., & Aronson, E. (2008). *Mistakes were made (but not by me): Why we justify foolish beliefs, bad decisions, and hurtful acts*. Houghton Mifflin Harcourt.

Taraban, R., Kerr, M., & Rynearson, K. (2004). Analytic and pragmatic factors in college students' metacognitive reading strategies. *Reading Psychology*, *25*(2), 67-81. https://doi.org/10.1080/02702710490435547

Teng, L. S., & Zhang, L. J. (2020). Empowering learners in the second/foreign language classroom: Can self-regulated learning strategies-based writing instruction make a difference? *Journal of Second Language Writing*, *48*, 1-16. https://doi.org/10.1016/j.jslw.2019.100701

Van Lier, L. (2008). Agency in the classroom. In J. P. Lantolf & M. E. Poehner (Eds.), *Sociocultural theory and the teaching of second languages*. Equinox.

Voge, N. (2011). Teaching developmental reading requires teaching students to read contexts as well as texts. *Research & Teaching in Developmental Education*, *27*(2), 82-86. https://www.jstor.org/stable/42802373

Willingham, D., & Price, D. (2014). Theory to practice vocabulary instruction in community college developmental education reading classes: What the research tells us. *Journal of College Reading & Learning*, *40*(1), 77-91. http://dx.doi.org/10.1080/10790195.2009.10850326

Winters, F., Greene, J., & Costich, C. (2008). Self-regulation of learning within computer-based learning environments: A critical analysis. *Educational Psychology Review, 20(4),* 429-444. http://dx.doi.org/10.1007/s10648-008-9080-9

Wolters, C. A., & Hussain, M. (2015). Investigating grit and its relations with college students' self-regulated learning and academic achievement. *Metacognition and Learning, 10*, 293-311. http://dx.doi.org/10.1007/s11409-014-9128-9

Wong, T. L., Xie, H., Zou, D., Wang, F. L., Tang, J. K. T., Kong, A., & Kwan, R. (2020). How to facilitate self-regulated learning? A case study on open educational resources. *Journal of Computers in Education*, *7*(1), 51-77.

Zhang, L. J., Aryadoust, V., & Zhang, D. (2016). Taking stock of the effects of strategies-based instruction on writing in Chinese and English in Singapore primary schools. In R. E. Silver & W. Bokhorst-Heng (Eds.), *Quadrilingual education in Singapore: Pedagogical innovation in language education* (pp. 103-126). Springer,

Zhang, L. J., Thomas, N., & Qin, T. L. (2019). Language learning strategy research in *System:* Looking back and looking forward, *System, 84*, 87-93. https://doi.org/10.1016/j.system.2019.06.002

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). Academic Press. http://dx.doi.org/10.1016/B978-012109890-2/50031-7

Zimmerman, B. J. (2013). From cognitive modeling to self-regulation: A social cognitive career path. *Educational Psychologist*, *48*(3), 135-147. https://doi.org /10. 1080 / 00 461520 2013.794676

Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology*, *82*(1), 51-59. https://doi.org/10.1037/0022-0663.82.1.51

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 299-315). Routledge.

Zimmerman, B., & Schunk, D. (2011). SRL and performance: An introduction and an overview. In B. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 187-202). Routledge.

Zimmerman, D. W., & Williams, R. H. (1998). Reliability of gain scores under realistic assumptions about properties of pre-test and post-test scores. *British Journal of Mathematical and Statistical Psychology, 51*(2), 343-351. https://doi.org/10.1111/j.2044-8317.1998.tb00685.x

Journal of Language Horizons

Scientefic Quarterly Journal of Language Horizons, *Alzahra University*
Volume 7, Issue 1, Spring 2023 (Serial No. 15)

ESL Writers' Performance in Exam and Non-Exam Academic Writing Settings

Research Article
pp. 59-85

Fahimeh Marefat[*1]

Mojtaba Heydari[2]

Received: 2021/08/29 Accepted: 2022/02/12

Abstract

With the growing access to new types of reference tools, today's L2 writers have a plethora of choices when completing an academic writing assignment. Such resources are absent in most high-stakes academic writing exams, making the two situations dissimilar. Aimed to compare the performances of ESL writers in Exam and Non-exam (real-life) academic writing situations, the present study recruited seven ESL university students who had previously taken an IELTS test. The students completed two analogous writing tasks: an exam-setting and a Non-exam writing test which aimed to simulate the real-life setting. Coh-Metrix analysis of the linguistic features of syntactic complexity, lexical sophistication, and text cohesion of the writings suggested that the students improved the textual quality of their writings in real-life academic writing situation. In addition, FACETS analysis of the quality of the writings, as assessed by the human raters, showed that the students did not benefit equally from the merits of the real-life settings compared to the Exam settings. The findings suggest that the students spent different amounts of time and used different types of queries to consult with external resources. Students' background training and writing strategies can highly affect their performance in real-life academic writing compared to the writing exams, warning against the validity of such tests.

Keywords: academic writing, Coh-Metrix, exam / non-exam settings, FACETS, writing exams validity

---

[*] Corresponding Author
[1] Professor, Allameh Tabataba'i University, Department of English Language and Literature, Faculty of Persian Literature and Foreign Languages, Tehran, Iran. fmarefat@atu.ac.ir. fahimehmarefat@yahoo.com
[2] Ph.D. Candidate, Allameh Tabataba'i University. Department of English Language and Literature, Faculty of Persian Literature and Foreign Languages, Tehran, Iran.Heydari_mojtaba@atu.ac.ir

DOI: 10.22051/lghor.2022.37489.1553
DOR: 20.1001.1.2588350.2023.7.1.3.1

Introduction

In the digital era, students' daily lives have been occupied by a variety of technology-facilitated practices such as web-browsing, texting, emailing, playing online games, and chatting through social network platforms (Zheng & Warschauer, 2017). In addition to their general life practices, students' academic lives heavily rely on digital resources as well: taking notes in the class, doing research, reading e-books, doing and submitting their course assignments, and being in contact with their teachers and peers. Regarding academic writing, a myriad of digital tools is available in assisting learners to compose their essays and enormously facilitating their process of writing. Unsurprisingly, today's students do most, if not the entire, of their academic writings using computers and, as a result, have a wider choice of tools to compose, edit, and enrich their writings (Yoon, 2016; Zhi & Huang, 2021).

Dictionaries and other traditional writing references are now available online and provide richer, faster, and easier access to information. (Dziemianko, 2012). Moreover, new types of online resources, like search engines, grammar checkers, corpus tools, and forums are now helping L2 writers to solve their lexical and grammatical problems (Yoon, 2016). Due to the significance of digital resources in learners' life, the ability to effectively utilize different reference resources should shape an essential part of digital literacy in academic settings (Conroy, 2010; Flowerdew, 2010; Kennedy & Miceli, 2010). Similarly, the use of online information is becoming an indispensable element of writing (Leijten et al., 2014). Therefore, as Hayes (2012) discusses, it should be included in future research to help achieve a better understanding of real-world L2 writing behaviors (Gánem-Gutiérrez & Gilmore, 2018).

However, a strong disconnect exists between students' classroom writing and exam writing contexts. While students are benefiting from the countless number of digital tools which assist them in the process of everyday classroom writing, the role of such resources is utterly overlooked in most high-stakes English exams. Today, while in many popular high-stakes academic writing tests such as IELTS (computer-delivered) and TOEFL, computers are the medium of text composition, the only digital asset at the writers' disposal is the text processor exclusively developed for composing the text. Therefore, students who have been practicing writing by getting assistance from digital resources, are deprived of such tools and

solely rely on their mental resources to write an essay.

Writing Processes

In the last two decades, the distinction between planning, composing, and revising has begun to erode entirely as a writer's craft has shifted from taking notes on papers to type and even dictate them (e.g., by Google, Siri, etc.) to be stored on their mobiles, laptops, or other digital companions. While the older models of writing emphasized the role of writer's memory in writing, recent models of writing additionally include the element of searching which recognizes the writer's use of external sources, like online dictionaries, to access information during writing (Leijten et al., 2014). The cognitive-based writing models consider writing as a problem-solving activity where the writers should approach the task as a problem and employ intellectual resources to solve it (Hyland, 2002). Flower and Hayes' (1981) model, for example, suggested that the writing process is recursive and involves planning, drafting, revising, and editing.

Inasmuch as writing processes have evolved in the last decades, the methods to observe and study these processes, too, have changed. L2 writing researchers have traditionally studied composition processes using direct observation of writers and their reflections and recounts on their writings to explore L2 writers' problem-solving processes and strategies (Bloom, 2008; Serror, 2013). With the emergence and development of digital technology, revolutionary observation methods including screen capture (e.g., Khuder & Harwood, 2015), keystroke logging (e.g., Serror, 2013), and eye-tracking (e.g., Gánem-Gutiérrez & Gilmore, 2018) have brought up more in-depth yet fairly quantifiable data to study writing composition processes.

Previous Studies

Despite the burgeoning use of digital resources, research on learner use of online resources has mainly been limited to individual reference resources in classroom settings (Yoon, 2016). Besides, studies comparing the writing processes and performances in exam and non-exam situations are scarce. Roca de Larios et al. (2008), for instance, investigated how different SL writers allocated time in the process of writing. They came up with two main conclusions: (i) the largest

percentage of composition time was spent on formulation and it was the predominant process for all groups and (ii) writers with different proficiency levels devoted different proportions of time to the writing processes; more skilled writers were more likely to regulate their composition processes.

In a recent study concerning the time of writing, Lee et al. (2021) found that when the L2 writers were given extra 15 minutes in a 30-minute argumentative writing test, the quality of argumentation was significantly higher in all the related subscores. In a very relevant study, Oh (2019) gave 39 English learners the chance to use extra resources to write an online review and compared their performance when they were not allowed to use the writing resources. He found that although the students performed better when they used resources during the writing process, their relative proficiency standings did not change. Oh concluded that giving access to the writing resources did not show any difference in distinguishing test-takers' levels.

Also, Khuder and Harwood (2015) investigated the product and process of writing in different situations. Ten graduate students wrote two argumentative essays under test and non-test conditions. The researchers used keystroke logging, screen recording, and stimulated recall protocols to observe their writing processes. They found statistically different time allocation for writing processes under the two settings. Besides, the participants received an average of .8 points higher for writing under the non-test condition. However, in Khuder and Harwood's study, the non-test situation failed to represent an authentic situation where students usually write the writing assignments. For instance, one of the researchers was present as the participants were writing under the Non-exam situation. In addition, as reported by the researchers, none of the participants, except for one, spent a significantly longer time in the non-test situation. This casts doubt on whether the students felt at ease, similar to a real-life situation, to spend as much time as they needed to complete their writing. Still more, authentic writing requires more than one sitting (Gánem-Gutiérrez & Gilmore, 2018).

This Study

As part of a bigger research project, this study set out to investigate how L2 writers navigate the growing variety of digital resources while completing an academic writing assignment and compare it against how they complete the task in a

situation similar, in many ways, to exam settings. This will provide insights as to how different learners can benefit from digital resources and how such affordances can reshape the future of L2 exams.

Research Questions

The present study sought the answer to the following research questions:

a) Is there any difference in the ESL writers' performance in terms of linguistic features, and overall, under the two academic writing settings of Exam vs. Non-exam?
b) Does the provision of an extended time for the writing process affect student's writing equally?
c) Does the provision of extra resources (e.g., dictionaries, wikis, etc.) affect student's writing equally?

Method

Participants

The call for participation was announced in five different ESL classes in a Canadian university with international students from a variety of language backgrounds. The participants were invited to take part in the research, being informed about the nature of the study and the time required to accomplish the tasks. From among the 83 respondents who expressed their initial willingness by filling out the online survey form, seven participants managed to finish the research procedure by attending all three phases of the study. Although this number of participants may be insufficient for quantitative studies, due to the qualitative nature of data collection and analysis in writing process studies (e.g., Khuder & Harwood, 2015) and the time and resources required to collect data for each participant, it seemed adequate for the purpose and scope of our study.

The students were from five language backgrounds (Persian, Spanish, Chinese, Arabic, and French) studying at graduate and undergraduate programs studying Engineering, Management, Linguistics, Philosophy, and Communication. At the time of the study, all the participants (four male and three were female students between 19 and 35) had an IELTS score (within two years of the time of the test). Also, three of them were preparing for a higher score. Table 1 shows the participants' biodata.

Table 1

The Participants' Biodata

Name	Sex	Age	L1	Major	School Year	IELTS Writing Band Score
Abdul	Male	35	Arabic	Linguistics	Post-graduate	7.5
Ali	Male	32	Persian	Management	Post-graduate	6.5
Lio	Male	19	Chinese	Engineering	Undergraduate	5.5
Pablo	Male	20	Spanish	Engineering	Undergraduate	6
Pari	Female	31	Persian	Engineering	Post-graduate	6.5
Sophie	Female	24	French	Philosophy	Graduate	7
Wei	Female	26	Chinese	Management	Graduate	6.5

Instruments

Computer Familiarity Questionnaire (CFQ). As suggested by previous research, the participants' familiarity with computers might have an impact on their performance in this study. Therefore, a CFQ was adapted from Weir et al.'s (2007) study, with a few adjustments for the present research context. The CFQ (see Appendix A) consists of 14 questions on a 5-point Likert Scale. A link to the online version of the questionnaire was sent to the email addresses of the participants. The CFQ took about 10 minutes on average for each participant to complete. The analysis of the results pointed out that the participant's familiarity with computers in terms of three aspects, i.e. Computer Usage, Comfort and Perceived Ability, and Interest in Computers were very high for all the participants as they all reported using computers very frequently at home and university.

The Writing Tasks. Two writing tasks were designed to collect student writings in the exam and real-life settings. In the Exam-setting, the participants were asked to write an essay in a 40-minute time limit on argumentative writing prompt with a general academic topic (Appendix B). While the writers used a computer to write their essays, they were only allowed to use Notepad to type their texts. This way, they didn't benefit from any spell/grammar-checkers, dictionaries, or thesauruses that might be enabled in a more sophisticated word processing software like MS Word. To simulate a high-stake writing exam setting, they were not allowed to use the internet, books, or any software during the 40-minute writing session.

In the Non-exam setting, the students were again asked to write an argumentative essay in response to a prompt similar to the one assigned to them in Exam-setting (Appendix B). However, in this task, the students were not given a 40-minute time limit. Instead, a 10-day deadline was provided to write their essays with no limitation in the hours or the number of sessions spent on the writing task within the deadline time frame. To make the task similar to the real-life academic writing tasks, the writers were given the choice to take our laptop equipped with the screen recording software to any preferred place (e.g., home, library), during the 10-day deadline to finish and submit their texts. During that time, they were also allowed to consult with any online resources including dictionaries, grammar-checkers, wikis, etc. To ensure maximum similarity between the two prompts, both prompts were selected from a pool of practice IELTS writing prompts provided by the British Council website. As for the comparability of the two prompts and the essays, two experienced IELTS examiners were consulted.

Screen Recording Software. To record the screen activities of the writers during both tasks, ApowerREC v 1.5 was installed on our laptop. The tool allows the users to easily save the screen-captured video and share it with others. We decided not to use keystroke logging since the writers spent some of their time using external resources such as Internet browsing which is not captured in keystroke logging software like Inputlog (Khuder and Harwood, 2015). Instead, we relied on the screen capture video to manually code the episodes of the writing, despite being more time-consuming, to capture and code all the elements of using external resources like searching and reading.

The Raters. Three raters were asked to blindly evaluate the essays written by the participants in the two settings. The raters were not informed that the texts were written by the same individuals under Exam and Non-exam settings. The raters were all IELTS teachers who had extensive experience in grading IELTS writing grading and were working as 'IELTS Mock Examiners' at language institutes. The raters were all males, at the ages of 33, 35, and 45 years old. The raters used the public version of the IELTS writing rubric to grade the tests and employed a 0-9 band score range based on the rubric.

Coh-Metrix Indices. One of the purposes of this study was to assess the degree to which linguistic features of the essays crafted in the Exam and Non-exam

settings differ. To compare the linguistic features of syntactic complexity, lexical sophistication, and text cohesion, the automatic computational linguistic tool of Coh-Metrix was employed (McNamara & Graesser, 2012).

Several indices were drawn from Coh-Metrix 3.0 to assess the quality of the essays across word, sentence, and discourse levels. The indices were selected based on the prior research and their application in the present study. For example, McNamara et al. (2010) found that syntactic complexity, lexical diversity, and word frequency were better predictors of essay quality. In the current study, we came up with 11 Coh-Metrix indices which proved valid in the literature and matched the IELTS writing assessment criteria. In addition, as accuracy is not measured by Coh-Metrix, we included one index to assess the accuracy of the essays. Also, one category was added to measure the length of words and sentences. The indices were categorized into five criteria and included:

a) *Length:* word count, word length, and sentence length
b) *Lexical Complexity:* word familiarity, lexical diversity, word frequency, content word concreteness
c) *Syntactic Complexity:* embedded clauses, number of modifiers per noun phrase, syntactic similarity, minimal edit distance
d) *Cohesion:* aspect repetition, content word overlap, connective incidences
e) *Accuracy:* error-free T-units divided by total T-units

Procedure

Phase 1. Exam-Setting. This phase aimed to simulate IELTS taks 2 writing tests. The writers were instructed to write no less than 250 words in 40 minutes in response to the selected prompt. The students completed this phase in a quiet university office using our equipped laptop. To collect the data for later analysis, ApowerREC software was used to record the screen activity of the writers from the beginning to the end of the task. At the end of the session, the writing document was saved in .txt format for later analysis.

Phase 2. Non-exam Setting. This phase was designed to simulate real-life academic writing tasks usually assigned for academic writing courses. As previously mentioned, the students were instructed to write on an argumentative writing prompt

within a 10-day deadline with access to external resources (e.g., dictionaries, grammar checkers, etc.). The participants took our laptop equipped with the screen recording and completed the task at the location of their choice. Before that, we guided them to run the screen-recording software before their writing activity. At the end of the 10-day deadline, the participants were asked to fetch the laptop where the final text and all the screen-recording data were stored.

Results

Processes

Using the data from our screen recordings and drawing upon Gánem-Gutiérrez and Gilmore's (2018) episode coding categories, we categorized the processes that the writers went through during the two tasks. Based on our modified version of categorization, there were five general episodes the writers underwent while writing:

- *Text construction:* Producing new text in the word processor
- *Re-reading:* Rereading their previously written text (as evident from eye-tracking data)
- *Revising:* Modifying the previously written text
- *Pausing:* Not being involved in any of the above-mentioned activities
- *Use of external resources (only in Non-exam setting):* Using external resources (e.g., online thesaurus, spell/grammar-checkers, etc.)

Using this coding scheme, one of the authors watched all the screen recordings, and the episodes were coded manually; the time spent on each episode was calculated. Finally, the sum of time spent on each episode was tabulated using MS Excel. Tables 2 and 3 show the duration of episodes for each process type in the Exam and Non-exam settings and their proportion regarding the overall time of the task.

Table 2

Duration of Episodes for Each Process Type in Exam Settings (Overall time for all the students was 40 minutes)

	Text construction	Re-reading	Revising	Pausing
Abud	47%	14%	21%	18%
Pari	43%	13%	20%	24%
Ali	35%	15%	19%	27%
Pablo	49%	12%	11%	28%
Wei	51%	9%	11%	29%
Lio	56%	10%	13%	21%
Sophie	48%	13%	23%	16%

Table 3

Proportional Duration of Episodes for Each Process Type in Non-exam Settings

	Text construction	Re-reading	Revising	Pausing	Use of External Resources	Overall time (minutes)
Abud	30%	17%	26%	28%	9%	92 minutes
Pari	33%	15%	20%	21%	11%	135 minutes
Ali	37%	11%	19%	27%	6%	85 minutes
Pablo	34%	16%	23%	23%	4%	73 minutes
Wei	39%	11%	20%	21%	9%	102 minutes
Lio	44%	9%	18%	18%	11%	94 minutes
Sophie	34%	16%	23%	15%	12%	128 minutes

As shown in Tables 2 and 3, Lio spent the highest time on text construction in both tasks while Ali and Abud spent the least in the Exam and Non-exam settings respectively. Also, for re-reading, Wei and Lio spent the lowest time in both tasks, and Ali and Abud spent the highest time in both tasks, respectively. An interesting observation is that Pablo spent the lowest time (4%) on using external resources while Sophie spent 12% of her time on it. This is interesting because they both spent a similar amount of their time on text construction, re-reading, and revising.

Table 4

Frequency and Types of the Queries for the Use of External Resources

Type of the Query	Abud	Pari	Ali	Pablo	Wei	Lio	Sophie
Thesaurus	4	6	2	0	2	3	8
Collocation	5	4	0	0	0	0	5
Translation	0	1	3	3	3	1	0
Grammar check	5	7	0	2	2	2	2
Preposition check	2	3	2	0	1	2	5
Writing Pattern	0	2	0	2	1	1	2
Word usage check	0	2	0	0	0	0	3
Grammar Study	0	1	0	3	1	0	0
Total	16	26	7	10	10	7	25

Another observation is the frequency and type of the queries carried out by the writers during the episode of using external resources. The queries were counted using the screen recordings and were categorized based on a modified version of Yoon's (2016) categorization. As shown in Table 4, Pari and Sophie had the highest, and Ali and Lio had the lowest number of queries compared to the other writers. Also, consulting with thesauruses and grammar checking are, respectively, the most common use of external resources among the writers.

Performances

Coh-Metrix Analysis. Coh-Metrix was used to analyze their underlying features of the collected essays in terms of the mentioned indices. The results from Coh-Metrix are summarized in Table 5.

Table 5

Coh-Metrix Indices for the Essays Written by the Participants in the Exam-setting and the Non-exam Setting

		Abud		Pari		Ali		Pablo		Wei		Lio		Sophie	
		Exam	Non-exam	Exam	Non-exam	Exam	Non-exam	Exam	Non-exam	Exam	Non-exam	Exam	Non-exam	Exam	Non-exam
Descriptive	Word Count	295	312	273	292	241	245	242	238	225	235	231	246	240	259
	Word Length	1.53	1.69	1.81	1.89	1.77	1.81	1.42	1.44	1.49	1.47	1.38	1.51	1.50	1.55
	Sentence Length	24.3	23.8	22.5	24.1	19.7	19.8	20.2	20.2	21.4	21.7	18.8	21.3	21.2	21.8
Lexical Complexity	Word Familiarity*	561	556.3	568.2	561.4	597.1	597	585.3	585.2	579	581.1	595.6	582.9	571.8	568
	Lexical diversity	0.83	.89	0.75	0.79	0.60	0.61	.63	0.65	0.60	0.61	0.58	0.61	0.70	0.74
	Word frequency*	2.09	2.01	2.16	2.01	2.43	2.43	2.37	2.35	2.39	2.41	2.34	2.27	2.30	2.21
	Content Words Concreteness*	346.6	343	333.5	339.6	357.1	356.2	361	360.5	354.8	352.1	368.1	359.2	353.6	349.9
Syntactic Complexity	Embedded Clauses	8.58	8.96	4.07	4.43	3.28	3.28	4.18	4.18	3.98	4.26	3.85	4.19	3.98	4.53
	Number of modifiers per noun phrase	1.04	1.25	0.91	0.96	0.86	0.89	0.83	0.82	0.85	0.87	0.82	0.88	0.91	0.97
	Syntactic Similarity*	0.089	0.082	0.086	0.071	0.085	0.085	0.084	0.086	0.087	0.085	0.083	0.085	0.070	0.081
	Minimal Edit Distance, part of speech	0.69	0.72	0.63	0.67	0.59	0.60	0.60	0.60	0.61	0.59	0.58	0.63	0.61	0.67
Cohesion	Tense and Aspect Repetition	0.72	0.71	0.74	0.76	0.69	0.71	0.68	0.67	0.71	0.70	0.66	0.71	0.68	0.69
	Content Word Overlap	0.034	0.029	0.049	0.042	0.075	0.070	0.057	0.051	0.054	0.049	0.057	0.048	0.047	0.040
	Connective Incidences	118.46	123.12	102.34	109.75	95.24	97.35	98.13	97.35	87.50	91.84	88.53	92.76	91.12	96.04
Accuracy	Error-free T-units divided by total T-units	0.95	1	0.92	0.97	0.78	0.81	0.91	0.91	0.87	0.90	0.81	0.93	0.82	0.89

To compare the indices, due to our limited sample size, non-parametric tests of Wilcoxon-Signed Rank and Chi-Square were run. The results are summarized as follows.

a) Length: For the first three textual indices, separate analyses of chi-square (only for the Word Count) and Wilcoxon-Signed Rank test for Word and Sentence

Lengths were conducted. Based on the results, although the ESL writers used more words in their writings in the Non-exam task (n = 1827, residual = 40, expected = 1787) compared to the exam task (n = 1747, residual = -40, expected = 1787), the results of the chi-square test (χ^2 (1) = 1.79, p > .05, Cramer's V = .022 representing a weak effect size) indicated that the difference was not significant. In terms of word length, the ESL writers had a slightly higher median score on the second task (Mdn = 1.55) than the first task (Mdn = 1.55). The results of Wilcoxon-Signed Ranked test (Z = -2.11, p < .05, r = .179 representing a weak effect size) indicated that there was a significant but weak difference between ESL writers' performance on the first and second tasks in terms of word length. Regarding the sentence length index, while the ESL writers had a slightly higher median score on the second task (Mdn = 21.70) than the first task (Mdn = 21.20), Wilcoxon-Signed Ranked test (Z = -1.57, p > .05, r = .143) showed that there was not any significant difference between ESL writers' performance on the first and second tasks.

b) Lexical Complexity: A non-parametric Wilcoxon-Signed Rank test was run to compare the ESL writer's performance in the first and second tasks in terms of lexical complexity. The ESL writers had a slightly higher median score on the Non-exam task (Mdn = 234.20) compared to the first task (Mdn = 234.06) in terms of lexical complexity. The results (Z = -2.36, p < .05, r = .250 representing a weak effect size) indicated that there was a significant but weak difference between ESL writers' performance on the first and second tasks in terms of lexical complexity.

C) Syntactic Complexity: Based on the results of a non-parametric Wilcoxon-Signed Rank test, the ESL writers had a slightly higher median score on the second task (Mdn = 1.45) than the first task (Mdn = 1.39). The results (Z = -2.20, p < .05, r = .214 representing a weak effect size) indicated that there was a significant but weak difference between ESL writers' performance on the first and second tasks in terms of syntactic complexity.

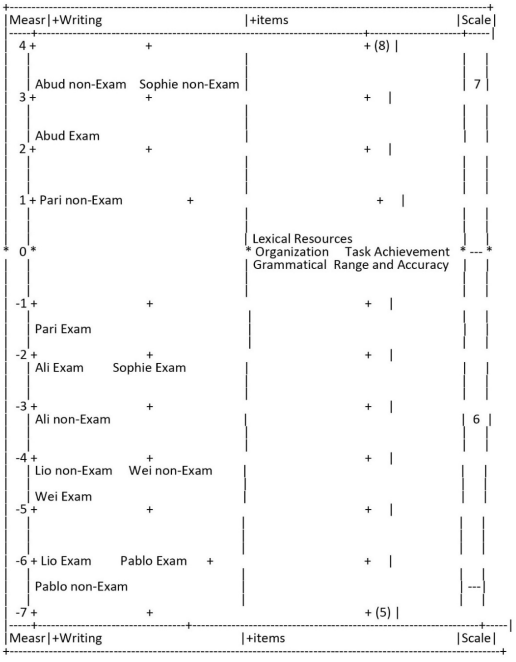
d) Cohesion: A non-parametric Wilcoxon-Signed Rank test showed that the ESL writers had a slightly higher median score on the second task (Mdn = 32.69) than the first task (Mdn = 32.00) in terms of cohesion indices. The results (Z = -2.19, p < .05, r = .179 representing a weak effect size) indicated that there was a significant but weak difference between ESL writers' performance on the first and second tasks in terms of cohesion.

e) Accuracy (error-free T-units): Another non-parametric Wilcoxon-Signed Rank test showed that the ESL writers had a slightly higher median score on the second task (Mdn = .91) than the first task (Mdn = .87) concerning error-free T-units. Based on the results ($Z = -2.21$, $p < .05$, $r = .214$ representing a weak effect size), there was a significant but weak difference between ESL writers' performance on the first and second tasks regarding error-free T-unit length.

Many-Facet Rasch Measurement. To compare the performances of the writers, the raw scores, on a 9-point scale, assigned to the essays by the raters were submitted to FACETS. In our model, we used three facets: raters (n=3), writings (seven students in two tasks, n=14), and the items (four criteria: Lexical Resources, Grammatical Range and Accuracy, Organization, and Task Achievement). However, for a cleaner representation, we do not report the measures related to the raters (after checking the inter-rater reliability). Fig.1 represents the relationship between the two facets of the model.

Fig. 1

The Relationship Between the two Facets of the Model



In this figure, technically referred to as Vertical Rulers, the left column is

the measurement ruler, labeled Measr. The values of this ruler are in logits, ranging from −7 to +4. The column labeled + Writing represents the quality of the writings, ranging from −6 to +3. This means that Abud Non-exam and Sophie Non-exam enjoyed the highest quality, and Pablo Non-exam showed the lowest quality among the writings. The column labeled + Items represents item difficulty based on the four rating criteria. It shows that students performed the best in the aspect of Lexical Resources, and the worst in terms of Grammatical Range and Accuracy, though the differences are very minimal, and their performance in Organization was on par with that of Task Achievement. Finally, the rightmost column shows the IELTS 9-point rating scale, ranging from 0 to 9.

Table 6

Student Writings Measurement Report

Rater	Total score	Logit	Error	Infit	Outfit
				MnSq	MnSq
Abud Non-exam	84	6.77	1.11	2.29	2.88
Sophie Non-exam	84	6.77	1.11	0.08	0.06
Abud Exam	83	5.69	0.95	0.74	0.44
Pari Non-exam	81	4.36	0.73	0.9	0.70
Pari Exam	74	0.70	0.84	0.89	0.48
Ali Exam	73	-0.09	0.95	0.64	0.3
Sophie Exam	73	-0.09	0.95	1.07	1.47
Ali Non-exam	72	-1.09	1.06	0.09	0.07
Wei Non-exam	71	-2.10	0.94	2.07	2.93
Wei Exam	70	-2.85	0.81	1.16	0.93
Lio Non-exam	70	-2.85	0.81	0.75	0.49
Pablo Exam	66	-4.90	0.69	0.76	0.68
Lio Exam	66	-4.90	0.69	1.13	1.23
Pablo Non-exam	65	-5.40	0.71	0.66	0.63

Table 7

Contrast Between Measurement Reports of Exam and Non-exam Writing for Each Student

	Rank in the Exam Task	Rank in the Non-exam Task	Ranking Change	Measure Contrast	t	P value
Abud	1st	1st	0	+0.98	0.70	0.24
Pari	2nd	3rd	-1	+2.64	1.88*	0.036
Ali	3rd	4th	-1	-0.93	0.67	0.25
Sophie	4th	2nd	+2	+4.90	3.52*	0.009
Wei	5th	5th	0	+0.68	0.58	0.28
Pablo	6th	7th	-1	-0.34	.50	0.31
Lio	7th	6th	+1	+1.93	1.65	0.056

** Significant difference at α=0.05(df=22)*

FACETS also produces detailed reports about the performance of individual writers in terms of total scores and logits (Table 6). The writings are ordered from the highest quality, on top, to the lowest quality, at the bottom of the table. Thus, Abud Non-exam and Sophie Non-exam were equally the best essays of all at +6.77 logits and a total score of 84. Moreover, Pablo Non-exam was showed the lowest quality at −5.40 logits and a total score of 65.

Table 7 shows the relative ranking of the students' writings in the two tasks and shows the changes in their standings from Exam to Non-exam. As shown in the table, Abud's writings were ranked first in both tasks and thus have shown no change in the rankings. The table also shows the contrast between the logit measures of the students' writing in both tasks. In Abud's case, the FACETS analysis shows +.98 of logits difference, meaning his Non-exam writing is better at 0.98 logits. To examine the significance of measure contrasts between the two tasks, separate paired t-tests are run using the following formula:

$$t = \frac{E \& NE\ Measure\ Contrast}{\sqrt{SET1^2 + SET2^2}}$$

Where E & N stand for Exam and Non-exam, and SET1 and SET2 stand for Standard Error of Measurement for Exam and Non-exam. Accordingly, the measure contrast between Abud's Exam writing and Non-exam writing is not significant (p = .24). Conversely, although Pari's writing in Non-exam has ranked

lower than her Exam writing, she has performed better in Non-exam at 2.64 logits which is statistically significant (p =.003). Ali's writing in Non-exam is also ranked lower than his writing in Exam, and the measure contrast of -0.93 shows a decrease in the quality of his writing, though the difference is not significant (p =.25). Sophie's writing, ranked 4th in the first task, stood 2nd in the second task with a high measure contrast of 4.90. This sharp increase in the quality of her writing (p = .009) explains Parisa's demotion in Non-exam writing despite its improved quality. Wei also performed slightly better in the second task at .68 logits (p = .28) but gained a similar standing in both tasks. Pablo, on the other hand, was demoted in the second task and had a slightly worse performance in the Non-exam setting (p = .31). Finally, Lio could promote his ranking over Pablo's in the Non-exam writing and showed a better performance compared to Exam writing at 1.93 logits, but the difference falls slightly lower than significant (p = .056).

Discussion

This study aimed to compare the performance of ESL students in the Exam writing situation with their real-life academic writing assignments. Based on the Coh-Metrix indices, the differences between the students' performance in the Exam and Non-exam tasks were significant but weak. This contradicts the findings of Riazi's (2016) study which showed similar performances for students doing two TOEFL writing tasks and one academic writing assignment. They reported similar indices for TOEFL and academic assignments on 15 measures of textual features including syntactic complexity (four measures), lexical sophistication (five measures), and cohesion (six measures). Riazi concluded that "the textual features of the texts produced in the test situation are not significantly different from those produced in the real-life academic writing" (p. 21). However, Riazi does not explicate the details of the process of real-life academic writing in his study. For example, the extent to which the students could or did consult the external resources remains unclear to us.

Based on the results, the writers generally performed better in Non-exam compared to Exam. An average of 0.4 increase from Exam to Non-exam in the IELTS scores given by the raters was observed for the students. This was less than what we expected according to Khuder and Harwood's (2015) 0.8 observed gain

from Exam to Non-exam writing. The results also suggested that the students did not benefit equally from the extra time and resources to improve the quality of their writings. As indicated by our findings in the FACETS analysis, while the overall quality of two students' writing in Non-exam significantly improved, three students' performance was enhanced only marginally, and two students underperformed in a real-life compared to the exam setting. Notably, the relative standing of the quality of the students' writing changed in the two tasks.

The difference in the gained score employing more time and external resources could be explained by the differences in the strategies the writers implemented to achieve their task goals. Given the stakes of exam situations, some students might employ specific strategies to gain a higher score in the exam which are not necessarily constructive for their real-life writing endeavors. Previous research has shown that students who are striving to get a high score on a test like IELTS, benefit from special coaching to improve their scores (Marefat & Heydari, 2018). Likewise, Pennycook (1996) reported that in the context of China the students were encouraged to practice writing on topics expected to appear in the test or memorize texts produced by renowned scholars, and use them in their writing when the topic is relevant. Such studies hint at cases where the applicants who are not really good at academic writing do very well in their tests. Similarly, in an interview conducted by Furneux (2013) one IELTS candidate mentioned that he was good in the IELTS exam but "rubbish" at real academic writing.

The analysis of the time spent by each student on different episodes of their writing as well as the types and frequencies of their queries on external resources bring to mind that the students who were successful in improving the quality of their writing spent more time on using external resources and carried out more queries during this episode. This is in agreement with Roca de Larios, et al. (2008)'s findings who concluded that the writers with different proficiency levels devote varying proportions of time to the writing processes, and the more skilled writers are more likely to regulate their composition processes. Also, similar to Khuder and Harwood's (2015) findings, it can be argued that the distribution of writing processes might have affected the quality of the essays.

A notion that can be taken into account in this regard is the notion of affordances (Hafner & Candlin, 2007; Yoon, 2016). From this perspective, what a

person does, depends on their abilities, goals, values, beliefs, and prior experience (Norman, 2013). An interesting study that draws on the notion of affordances in writing is Hafner and Candlin's (2007) study that specifically monitored the use of a language reference tool to improve L2 writing. The researchers observed the use of a specialized corpus by law students to support their legal writing assignments. The apprentice lawyers, their study implied, used the corpus mostly for legal document searches rather than for lexical or grammar patterns. Although the corpus was provided for them to help them with their word choice and grammar, the tendency of the writer to use it for legal support showed how their identity as lawyers and their membership of that culture strongly influenced their affordance. It should also be considered that due to the difference between real-life and test situations, the writers may tend to have a different conceptualization of the task. Curry (2004), for example, found that in the test situation, writers focus on the word level (grammar and vocabulary choice) rather than idea generation and argumentation as they do not have any resources. However, the writers can be more focused on such aspects of their writing when the material they require is available.

In a longitudinal study of six ESL writers' web-accessed corpus tool use, Yoon (2008) found that the frequency and range of corpus consultation, the types of strategies, and analyses employed by the participants were mostly determined by multiple factors ranging from individuals' prior experiences to disciplinary characteristics. One valuable insight from the study is that learners' motivation to use the corpus technology is determined by the extent to which they have meaningful engagement with it in the process of performing their real-life writing tasks.

The findings of our study corroborate the findings of Yoon (2008) which revealed that the participants' attitudes toward using reference resources as writing assistance were widely different. The cross-case analysis conducted in their study showed that the differences were mainly due to the multifaceted interactions of factors related to the text, writer, and context. This is in line with the implications of Ho Yung and Cai's (2020) study who discussed that writers with high English proficiency do not necessarily perform better in the real-life academic situation and their real-life performance depends on a plethora of factors.

A clear implication of this study would be the long-term impact of

students' past experience and training on their future real-life requirements. Understanding such individual differences and their roots can significantly inform learner training (Kormos, 2012) in the use of online reference tools, which has largely been lacking in the L2 writing pedagogy. In the same vein, in a case study of three Italian L2 writers revising their own compositions while consulting a corpus of Italian texts, Kennedy and Miceli (2010) showed how individual learners' attitudes, goals, and computer literacy affected their ability and willingness to use the unique functions of the corpus. The researchers concluded that the functions of corpus consultation should be explicitly taught.

Another potential implication of this study centers around the validity of conventional timed-impromptu writing tests. The observed gap in processes and performances of students between the exam and real-life academic settings raises serious concerns about the validity of such tests. Twenty-first-century students are now digital natives and are at ease with technology and, therefore, using computers and online resources as sources of assistance has now become a norm for them. As a result, students taking writing tests must perceive the relevance of their test experience to their current and future experiences of writing at university (Chan et al., 2017). In addition, it has been long believed that the restrictions imposed on the writers affect all of them equally and as Worden (2009) mentions, such tests are usually assessed with a 'lower' bar to make it fair for the students. However, as observed in this study, the students might not enjoy equally from lifting the exam restrictions as some of them are already coached for the exam situations and not the real-life venues. Therefore, it is an oversimplification to ignore inherent differences in the psychological characteristics of the writers and the potential differences in the strategies and styles they employ during the complex process of writing. This suggests the need for implementing practical solutions to accommodate the tools like corpora and dictionaries.

The interpretation of the results of this study points to the need for training learners to use technology which should consist of both initial and ongoing scaffolding to implement resources in their L2 writing (Yoon, 2016). In this regard, while the students are taught the shortcuts to improve the quality of their writings in exam settings, they should also be taught to use more general strategies to improve the quality of their writing in real-life tasks.

The findings should, however, be interpreted within the limitations of the study. The first limitation of the study is the number of participants. Due to the highly demanding and time-consuming nature of the tasks involved in this study, despite our attempt to recruit more participants, we were finally able to recruit seven participants. Therefore, the conclusions drawn based on the findings of this study should be considered cautiously. In addition, the weak effect sizes in the Coh-Metrix data analysis and the descriptive nature of some of our observations (e.g., time and frequencies of queries made on using external resources) prevent making strong generalizability claims.

Another complication of the study was to simulate a real-life academic writing setting. Despite the attempts made in the design and execution of the study, we believe asking the writers to record their screen whenever they wanted to write might interfere with the natural process of their writing. Having said that, some students might try to impress the raters while others might not be motivated enough to devote as much time and energy they do in reality. As pointed out by previous studies, when the writers are not responding to an authentic test, they might not be motivated enough (Khuder & Harwood, 2015).

Future research might replicate the current study with bigger sample sizes. In doing so, it is important to compare the students at identical proficiency levels but with different pedagogical backgrounds to eliminate the proficiency variable. Finally, conducting qualitative studies where students are observed in longer intervals and thicker data are obtained from students' writing processes will certainly be revealin

References

Bloom, M. (2008). Second language composition in independent settings: Supporting the writing process with cognitive strategies. In S. Hurd & T. Lewis (Eds.), *Language learning strategies in independent settings* (pp. 103-118). Multilingual Matters.

Chan, S., Bax, S., & Weir, C. (2017). Researching participants taking IELTS Academic Writing Task 2 (AWT2) in paper mode and in computer mode in terms of score equivalence, cognitive validity and other factors. *IELTS Research Reports Online Series, No. 4*. British Council, Cambridge English Language Assessment and IDP: IELTS Australia.

Conroy, M. A. (2010). Internet tools for language learning: University students taking control of their writing. *Australasian Journal of Educational Technology*, *26*(6), 861-882. https://doi.org/10.14742/ajet.1047

Curry, M. J. (2004). UCLA community college review: Academic literacy for English language learners. *Community College Review, 32*(2), 51-68. https://doi.org/10.1177/009155210403200204

Dziemianko, A. (2012). On the use(fulness) of paper and electronic dictionaries. In S. Granger & M. Paquot (Eds.), *Electronic lexicography* (pp. 319-341). Oxford University Press.

Flower, L. S., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*(4), 365-387. https://www.doi.org/10.2307/356600

Flowerdew, L. (2010). Using corpora for writing instruction. In A. O'Keeffe, & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics (*pp. 444-457). Routledge.

Furneux, C. (2013). What are the academic writing requirements of Masters level study in the Humanities and how far can EAP proficiency tests, such as IELTS, replicate them? Paper presented at the *CRELLA Summer Research Seminar*, United Kingdom.

Gánem-Gutiérrez, G. A., & Gilmore, A. (2018). Tracking the real-time evolution of a writing event: Second language writers at different proficiency levels. *Language Learning*, *68*(2), 469-506.

Hafner, C. A., & Candlin, C. N. (2007). Corpus tools as an affordance to learning in professional legal education. *Journal of English for Academic Purposes, 6*(4), 303-318. https://doi.org/10.1016/j.jeap.2007.09.005

Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication*, *29*(3), 369-388. https://doi.org/10.1177/0741088312451260

Ho Yung, k. W., & Cai, Y. (2020). Do secondary school-leaving English examination results predict university students' academic writing performance? A latent profile

analysis. *Assessment & Evaluation in Higher Education*, *45*(4), 629-642. https://doi.org/10.1080/02602938.2019.1680951

Hyland, K. (2002). *Teaching and researching writing*. Longman.

Kennedy, C., & Miceli, T. (2010). Corpus assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning and Technology*, *14*, 28-44.

Khuder, B., & Harwood, N. (2015). L2 writing in test and non-test situations: Process and product. *Journal of Writing Research, 6*(9)*,* 233-278. https://doi.org/10.17239/jowr-2015.06.03.2

Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing*, *21*(4), 390-403. https://doi.org/10.1016/j.jslw.2012.09.003

Lee, S., Lim, G. S., & Basse, R. (2021). The effect of additional time on the quality of argumentation in L2 writing assessment: A mixed-methods study. *Language Assessment Quarterly, 18*(3), 253-272. https://doi.org/10.1080/15434303.2021.1872080

Leijten, M., Van Waes, L., Schriver, K., & Hayes, J. R. (2014). Writing in the workplace: Constructing documents using multiple digital sources. *Journal of Writing Research*, *5*(3), 285-337. https://doi.org/10.17239/jowr-2014.05.03.3

Marefat, F., & Heydari, M. (2018). English writing assessment in the context of Iran: The double life of Iranian test-takers. In T. Ruecker & D. Crusan (Eds.), *The politics of English Second Language writing assessment in global contexts* (pp. 67-71). Routledge.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). The linguistic features of quality writing. *Written Communication, 27*(1), 57-86. https://doi.org/10.1177/0741088309351547

McNamara, D. S., & Graesser, A. C. (2012). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In P. M. McCarthy & C. Boonthum-Denecke (Eds.), *Applied natural language processing: Identification, investigation, and resolution* (pp. 188-205). IGI Global.

Norman, D. A. (2013). *The design of everyday things: Revised and expanded edition*. Basic Books.

Oh, S. (2019). Second language learners' use of writing resources in writing assessment, *Language Assessment Quarterly*, *17*(1), 60-84. https://doi.org/10.1080/15434303.2019.1674854

Pennycook, A. (1996). Borrowing others' words: Text, ownership, memory, and plagiarism. *TESOL Quarterly, 30*(2), 201-230. https://doi.org/10.2307/3588141

Riazi, A. M. (2016). Comparing writing performance in TOEFL-iBT and academic assignments: An exploration of textual features. *Assessing Writing*, *28*, 15-27. https://doi.org/10.1016/j.asw.2016.02.001

Roca de Larios, J., Manchón, R., Murphy, L., & Marín, J. (2008). The foreign language writer's strategic behaviour in the allocation of time to writing processes. *Journal of Second Language Writing, 17*(1), 30-47. https://doi.org/10.1016/j.jslw.2007.08.005

Serror, J. (2013). Screen capture technology: A digital window into students' writing processes. *Canadian Journal of Learning and Technology, 39*(3), 1-16. https://doi.org/10.21432/T28G6K

Weir, C. J., O'Sullivan, B., Yan, J., & Bax, S. (2007). Does the computer make a difference? Reaction of participants to a computer-based versus a traditional handwritten form of the IELTS writing component: Effects and impact. *IELTS Research Report*, *7*(6), 1-37.

Worden, D. (2009). Finding process in product pre-writing and revision in timed essay responses. *Assessing Writing, 14*(3), 157-177. https://doi.org/10.1016/j.asw.2009.09.003

Yoon, C. (2016). Individual differences in online reference resource consultation: Case studies of Korean ESL graduate writers. *Journal of Second Language Writing*, *32*, 67-80. https://doi.org/10.1016/j.jslw.2016.04.002

Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning & Technology, 12*(2), 31-48.

Zheng, B., & Warschauer, M. (2017). Epilogue: Second language writing in the age of computer-mediated communication. *Journal of Second Language Writing*, *36*, 61-67. https://doi.org/10.1016/j.jslw.2017.05.014

Zhi, M., & Huang, B. (2021). Investigating the authenticity of computer- and paper-based ESL writing tests. *Assessing Writing*, *50*, 100548. https://doi.org/10.1016/j.asw.2021.100548

Appendix A

Computer Familiarity Questionnaire

(Adapted from Weir et al., 2007)

Name:_____ (Given name) _____ (Surname)

Student number:_____ Gender: M/F (Please circle)

IELTS result (latest): Overall band:_____ Writing band:_____

There are 15 questions about your experiences and opinions of working with computer. Please **CIRCLE** the number to indicate your answer to each question/statement.

		Always	Very often	Often	Once a while	Never
1)	How often is there a computer available to you to use at each of the following places?					
	a) home	5	4	3	2	1
	b) University	5	4	3	2	1
	c) Public library	5	4	3	2	1
	d) Others (please specify): _____	5	4	3	2	1
2)	How often do you use a computer at each of the following places?					
	a) home	5	4	3	2	1
	b) University	5	4	3	2	1
	c) public library	5	4	3	2	1
	d) others (please specify): _____	5	4	3	2	1
3)	How often do you use computer for					
	a) surfing the Internet ?	5	4	3	2	1
	b) electronic communication, e.g. emails?	5	4	3	2	1
	c) study-related activities?	5	4	3	2	1
	d) Others (please specify): _____	5	4	3	2	1
4)	How often do you use the computer **software** for					
	a) games?	5	4	3	2	1
	b) word processing?	5	4	3	2	1
	c) spreadsheets?	5	4	3	2	1
	d) painting or graphics?	5	4	3	2	1
	e) data or text analysis?	5	4	3	2	1
	e) Others (please specify): _____?	5	4	3	2	1

4)	How often do you use the computer **software** for					
	a) games?	5	4	3	2	1
	b) word processing?	5	4	3	2	1
	c) spreadsheets?	5	4	3	2	1
	d) painting or graphics?	5	4	3	2	1
	e) data or text analysis?	5	4	3	2	1
	e) Others (please specify): _____?	5	4	3	2	1
5)	How often do you take a test on					
	a) paper?	5	4	3	2	1
	b) computer?	5	4	3	2	1
		Very comfortable	Quite Comfortable	Comfortable	Quite uncomfortable	Very uncomfortable
6)	How comfortable are you with using a computer in general?	5	4	3	2	1
7)	How comfortable are you with using a computer to write a paper?	5	4	3	2	1
8)	How comfortable are you with taking a test on					
	a) computer?	5	4	3	2	1
	b) paper?	5	4	3	2	1
9)	How do you feel about using the keyboard (typing)?	5	4	3	2	1
10)	It is very important to me to work with a computer.	5	4	3	2	1
11)	To play or work with a computer is really fun.	5	4	3	2	1
12)	I use a computer because I am very interested in this.	5	4	3	2	1
13)	I forget the time, when I am working with the computer.	5	4	3	2	1
		Excellent	Good	Fair	Poor	Very poor
14)	If you compare yourself with other students, how would you rate your ability to use a computer?	5	4	3	2	1

Appendix B

Writing Task Prompts

Writing Task 1 prompt

Environmental pollution has become so serious that many countries are trying to solve them. What are the most serious problems associated with it and what solutions can you suggest?

Writing Task 2 prompt

The internet has transformed the way information is shared and consumed, but it has also created problems that did not exist before. What are the most serious problems associated with the internet and what solutions can you suggest?

Journal of Language Horizons

Scientefic Quarterly Journal of Language Horizons, *Alzahra University*
Volume 7, Issue 1, Spring 2023 (Serial No. 15)

Ideology and the Sublime in William Wordsworth's *The Prelude*

Research Article
pp. 87-106

Ghiasuddin Alizadeh[*1]
Masoud Farahmandfar[2]

Received: 2021/07/20 Accepted: 2022/02/12

Abstract

In order to follow the contours of Wordsworth's ideological development and the role played by the sublime in the growth of his mind, one is well-advised to follow the temporal and narrative progression of *The Prelude* (1799). Through a Žižekian analysis of Wordsworth's *magnum opus,* this study argues that the sublime can be approached from two different points. First, it can be conceived as the symbolic intrusion into the imaginary world of the subject, rifting the sense of unity and perfection, and relocating the object of desire from the imaginary to the symbolic order. Next, it can be perceived as the intrusion of the real into the symbolic order, as the lack at the heart of the Other, making a perfect correspondence of the signifying elements and the ultimate achievement of the object of desire impossible. What links these two different significations of the sublime together is the crucial realization that, in both cases, the sublime holds the place of illusory and non-existent states.

Keywords: Wordsworth, *The Prelude*, sublime, Žižek, Romantic literature

[*] Corresponding Author
[1] Assistant Professor, Department of English Language and Literature, Faculty of Literature and Human Sciences, Malayer University, Malayer, Iran.
Ghiasuddin.alizadeh@malayeru.ac.ir
[2] Assistant Professor, Department of English Language and Literature, Faculty of Persian Literature and Foreign Languages, Allameh Tabataba'i University, Tehran, Iran. farahmand@atu.ac.ir

DOI: 10.22051/lghor.2022.36979.1528
DOR: 20.1001.1.2588350.2023.7.1.4.2

Introduction

Like many of his contemporaries, William Wordsworth (1770-1850) became intellectually involved with the sublime— that power in nature and art which inspires awe and deep emotion and which is manifest in grand and wild natural scenes (Day, 2001). One of the main literary sources in which Wordsworth put forth his theory of the sublime is *The Prelude* (1799), the autobiographical account of the poet's life and the growth of his intellectual and imaginative powers. In order to gain an insight into Wordsworth's opinion about the ongoing philosophical debate regarding the sublime, therefore, we have to rely on his poetry and try to disentangle his theoretical viewpoints from the warp and woof of *The Prelude*. The amount of his dependence on, or his deviation from, the popular discussions on the concept can be measured only after a careful analysis of his colossal autobiography.

In what follows, we will try to show that the sublime played a more drastic role in Wordsworth's life than a mere abstract and philosophical concept which would have no significant bearing on the subject's social reality. Drawing on the theories of Slavoj Žižek (1949), we will argue that the notion of the sublime provided the post-revolutionary Wordsworth with the opportunity to justify his failure in actualizing his potential artistic imagination, as well as the failure of the French Revolution in realizing the apocalyptic prophesies attributed to it by the supporters of freedom and liberty all across the Europe. The sublime object(s) of Wordsworth's ideology, which constitute the focal points of *The Prelude*, point to the desperate effort of the poet to salvage a foothold for himself in a world where the long-expected angel of liberty turned out to be nothing but a devil spreading its grim shadow over the expanse of the universe.

Literature Review

Numerous studies have been carried out on the subject of the sublime in Wordsworth's poetry such as Heffernan (1967), Owen (1973), Wlecke (1973), Kelley (1984), Stoddard (1985), Myers (2008). These have mostly, if not exclusively, focused on Wordsworth's theoretical direction, trying to trace either similarities or dissimilarities with the major philosophical and psychological arguments of the time. Yet, few, if any, have ever dealt with the ideological effects

of the sublime on the poet's life and thought. For instance, two of the most significant scholarly studies in the field, Hartman (1964) and Weiskel (1976), cover the philosophical and psychological aspects of the work. For Hartman, the sublime brings the poet face to face with the ontological incapacity of Nature in meeting the needs and desires of imagination; it is a reminder that natural environment, regardless of all its sublimity, cannot satisfy the insatiable appetite of the mental faculty. Weiskel, on a different plane, links the concept of the sublime to the Freudian theory of the Oedipal complex, as "the very moment in which the mind turns within and performs its identification with reason" (pp. 93-4).

In his groundbreaking study of the Romantic poetry, *Natural Supernaturalism,* M. H. Abrams (1973) argues that Wordsworth, in line with the major literary and intellectual figures of the period, believed in the possibility of attaining self-consciousness as the ultimate purpose behind human existence. Freedom, Abrams contends, was for the Romantics the final moment in the progress towards the completion of the historical process, and it was the French Revolution, with its emphasis on freedom and liberty, that kindled in the heart of these intellectuals the hope that the time had finally arrived for humanity to build the prophesied Heaven on earth. Even when the Revolution failed in realizing its dreams, Abrams claims, the Romantics did not give up their hope of a free world; rather, they came to the conclusion that Revolution should begin, first of all, in the mind of the individual if it is to appear on the social and political scale. Thus, to Abrams, *The Prelude* is the account of the fall of the human mind from unity into discord only to be restored to a higher unity after the cyclical progress of the soul is complete and the mind is reconciled to what it believed to be impediments on its way toward self-consciousness.

Richard Stang (1966) follows Abrams' steps in his study of *The Prelude*. The poem, Stang argues, opens in a sense of frustration and mental agitation, since the poet (i.e., the voice) states that he does not have the power to embark on the colossal project of writing a poem of an epic scale. This initial moment of disappointment, Stang maintains, is the result of the fact that the poet looks for the outside world as a guide to help him through the twists and turns of his literary odyssey. However, as soon as the focus shifts from the outside to the inside, that is to say, as soon as the unconscious replaces the conscious mind as the source of

inspiration, composition begins and, as Stang concludes *a la* Abrams, "the final failure of purely political, i.e., external liberty" turns out to be nothing but "a false dawn" (p. 65), which will necessarily be succeeded by the true arrival of the morning.

Brooke Hopkins's (1994) article focuses on the ideological overtones of *The Prelude*. Drawing on the theories of Mikhail Bakhtin (1895-1975), Hopkins argues that the poem is a polyphonous construct in which different voices enter into a dialogical relationship, each pertaining to a specific stage in the life of the poet. This simultaneous co-presence of the voices adds to the ambiguity of the poem, as well as giving the poet the opportunity to carry out the project of self-criticism as he gives vent to his different ideological viewpoints throughout his life. Hopkins concludes that the poem denounces the illusory nature of the poet's previous conceptions of freedom in favor of a "genuine liberty" which "is meant to be attained through the labor of its [i.e., the poem's] reading" (p. 299).

Nonetheless, whether Wordsworth was a Kantian in his distinguishing between the phenomenal and the noumenal worlds, seeing the sublime as the manifestation of the latter through a failure of the former, or he followed Burke in opposing the sublime to the beautiful and seeing in it a remedy for the weakness of the mind and soul caused by being too much exposed to beautiful shapes and scenes, it does not give us any clue as to the efficacy of the sublime in the poet's ideological and philosophical development. It seems to be the proper time to abstract from the arguments about the nature of the sublime and its aesthetic effects in poetry and prose, and focus, instead, on the role it plays in the functioning of ideology and the constitution of the ideological subject.

Methodology and Approach

In *The Sublime Object of Ideology,* Žižek puts forth a painstaking and hairsplitting analysis of the Marxist concept of ideology by drawing the attention to the various ways in which an ideological discourse succeeds in deceiving ideological subjects into behaving as if it were possible to escape the ontological deadlocks and antagonisms which inherently mark the socio-symbolic order. Apropos of the problematic of ideology in relation to reality and fantasy, Žižek (2008b) maintains:

> *Ideology is not a dreamlike illusion that we build to escape insupportable reality; in its basic dimension it is a fantasy-construction which serves as a support for our 'reality' itself: an 'illusion' which structures our effective, real social relations and thereby masks some insupportable, real, impossible kernel… The function of ideology is not to offer us a point of escape from our reality but to offers us the social reality itself as an escape from some traumatic, real kernel. (p. 45)*

Drawing on the Kantian notion of the sublime, Žižek equates the function of ideological fantasy with the function of the transcendental categories which, according to Kant, make possible the phenomenological appreciation of the noumenal Thing-in-Itself. "The way fantasy functions," Žižek (2008b) argues, "can be explained through reference to Kant's *Critique of Pure Reason*: the role of fantasy in the economy of desire is homologous to that of transcendental schematism in the process of knowledge" (p. 133). In other words, like the Kantian notion of the sublime, ideology provides the subject with the illusion that beyond the ontological loss, beyond the social and political antagonism, there exists a consistent and unified concept of society which one day will turn into reality. "According to Lacan, a sublime object is an ordinary, everyday object which, quite by chance, finds itself occupying the place of what he calls *das Ding*, the impossible-real object of desire" (Žižek, 2008b, p. 221). In contrast to the vulgar Marxist conception of ideology as the distortion of reality, Žižek perceives ideology as a fantasy construct which, through externalizing the inconsistency inherent to the symbolic order as such, constitutes the very notion of reality. "Ideology is not simply a 'false consciousness', an illusory representation of reality, it is rather this reality itself which is already to be conceived as 'ideological'" (Žižek, 2008b, p. 15). That is to say, beyond the limits of ideology, there is no reality; rather, the only thing which awaits the subject beyond the veil of ideology is the empty place of the Thing.

In the same vein, Žižek (2006) probes into the nature of what he calls the "parallax gap" which separates the modes of thinking which can never become reconciled through a consistent and harmonious synthesis: "the apparent displacement of an object (the shift of its position against a background), caused by a change in observational position that provides a new line of sight" (p. 17). Žižek

(2006) explains this parallax gap as it functions in the three discursive domains of philosophy, science and politics. For him, such a parallax gap constitutes the very reality in which we live:

> *First, there is the ontological difference itself as the ultimate parallax which conditions our very access to reality; then there is the scientific parallax, the irreducible gap between the phenomenal experience of reality and its scientific account/explanation… last, but not least, there is the political parallax, the social antagonism which allows for no common ground between the conflicting agents. (p. 10).*

That is to say, our socio-symbolic reality is overladen with gaps which can never be filled, with conflicts and antagonisms which can never be resolved nor reconciled, since there can never be a "common ground between the conflicting agents." Reality, Žižek contends, is a bundle of contingent and chaotic events between which no synthesis can ever be established. What we should do, as critics of ideology, is to forsake the notion of reality as a unified and consistent whole in favor of a notion of reality as rent with parallax gaps and unresolvable conflicts. That is to say, "if we look at the element which holds together the ideological edifice…from the right…perspective, we are able to recognize in it the embodiment of a lack, of a chasm of non-sense gaping in the midst of ideological meaning" (Žižek, 2008b, p. 110).

Žižek's ideas about the fundamental role of the sublime in the constitution and the survival of an ideological field can be very helpful in dismantling the ideological layers of Wordsworth's *The Prelude* and in finding convincing answers to questions which have plagued the critical literature ever since the poem's publication.

Argument and Discussion

The Prelude opens with the poet's description of an autumnal day, when he has recently returned from the dull and tiresome life in London to the Lake District, the beloved countryside where he spent his childhood and adolescence. The prospect of living in natural surroundings, free from the cumbersome limitations and inconveniencies of civilization, kindles in Wordsworth's heart a flame of joy and

happiness, and brings him the tidings that poetic genius can be once again awakened from the long sleep of lethargy:

The earth is all before me: with a heart
Joyous, nor scared at its own liberty,
I look about, and should the guide I chuse
Be nothing better than a wandering cloud,
I cannot miss my way. (Book I, ll. pp. 15-19)

From this very early stage, a careful, analytic reading reveals a problematic, hidden, meaning behind the apparent signification of the words and sentences. It seems that the poet is disturbed by his consciousness of an embarrassing state of things, by his recognition that there is something wrong with the account he presents to the reader. In *Fundamentals of Psychoanalytic Technique* (2007), Bruce Fink points to the processes of "unprovoked denial" and "overemphasized assertion" as a way to reach the true meaning behind the analysand's words:

> *In this form of denial [i.e. unprovoked denial], the analysand insists that something is not the case even when no one has claimed that it is…. In such cases of unprovoked denial, one can always ask why someone is taking the time and energy to deny something that no one in the context at hand…has in any way suggested or affirmed. One could retort that…the analysand is simply trying to forestall a conclusion that he assumes the analyst will jump to. Similar to unprovoked denials are what I call "overemphasized assertions." Here the analysand (or politician, business leader, or someone else) affirms something so forcibly and repeatedly that the listener begins to wonder why: if the speaker so fervently believes what he is saying, why does he feel the need to stress it so appreciably? (pp. 41-42)*

The subject does all he can in order to prevent the addressee from getting too close to his unconscious truth, to what he really is behind the mask of appearances. Moreover, in order to conceal the lack in himself, the subject creates the illusion of wholeness and completeness for the other's gaze, putting the onus of probable failure on the shoulders of external, adverse elements. Perhaps the most noticeable feature of *The Prelude* is the highly emphatic language which

Wordsworth employs in order to assure the reader of his intimate relationship with Nature, of the fact that he *cannot* fail because he has been under the tutorship of Divinity throughout his life. On the other hand, the repetitive negation of certain issues points to a probable (albeit, perhaps unconscious) attempt on behalf of the poet to conceal a traumatic knowledge. For instance, we should in no way take lines such as "with a heart / Joyous, nor scared at its own liberty," "I cannot miss my way," "not mine, and such as were not made for me" as neutral informative sentences; rather, we should take them as highly biased and pointed speech which would reveal a significant truth about the poet's life. It seems, at this early stage, that *the whole poem is composed in order to account for the failure of Wordsworth's ideological life as such*.

The breeze which blows on the poet arouses within him "a correspondent breeze," which slowly turns into a powerful tempest and breaks the "long-continued frost" of imagination. Wordsworth's heart is filled with the hope that finally, having been delivered from the bonds of his "unnatural self," the wells of his imagination will brim with water once again, and he will be able to carry out his life-long project of writing a monumental and everlasting work. In order to follow such a high and noble purpose, the lines continue, the poet lacks none of the prerequisites. He has a "vital Soul" and a "living Mind," which are the "first great gifts" bestowed on the great masters of poetry; his mind is replete with "general Truths", the "Elements and Agents" of the poetic mind; in one word, he is in possession of everything "needful to build up a Poet's praise." Also, there are numerous themes on which his mind can settle, from tales of love and chivalry, and stories of war for religion and liberty, to personal epics and philosophic songs. However:

> *... I have been discouraged; gleams of light*
> *Flash often from the East, then disappear*
> *And mock me with a sky that ripens not*
> *Into a steady morning: if my mind,*
> *Remembering the sweet promise of the past,*
> *Would gladly grapple with some noble theme,*
> *Vain is her wish; where'er she turns she finds*
> *Impediments from day to day renewed. (Book I, ll. pp. 134-141)*

Wordsworth's bitter confession about his constant failure to fulfill his

hopes and dreams points to a highly significant fact regarding the ideological nature of *The Prelude*. One of the main tenets of any ideological belief is the ability to invert the cause of its failure from the inside to the outside world. (Žižek, 2008b, p. 142). That is to say, when the symbolic reality fails to achieve its professed goals and aims due to purely internal problems and shortcomings, ideology locates the roots of this failure in some external ground, projecting the cause of failure to some foreign object or element. Of course, this constitutes the negative aspect of the sublime object of ideology, as "an object which is just an embodiment of the lack in the Other, in the symbolic order" (Žižek, 2008b, p. 192). Such an object finds a pivotal role in the ideological discourse since it is through the materialization of failure in this object that ideology succeeds in keeping the social and symbolic façade from falling and shattering into pieces. Wordsworth, as a subject of ideology, fails to perceive the 'true' cause of his constant failure to fulfill his dreams. Wherever he turns, he finds impediments "from day to day renewed," impediments which he believes prevent him from the germination of his poetical and imaginative faculties:

...Thus my days are passed
I live, a mockery of the brotherhood
Of vice and virtue, with no skill to part
Vague longing that is bred by want of power,
From paramount impulse not to be withstood,
A timorous capacity, from prudence;
From circumspection, infinite delay.
Humility and modest awe themselves
Betray me, serving often for a cloak
To a more subtle selfishness, that now
Doth lock my functions up in blank reserve,
Now dupes me by an over-anxious eye
That with a false activity beats off
Simplicity and self-presented truth. (Book I, ll. pp. 238-251)

Here, Wordsworth mentions in passing all the reasons which he believes cause his never-ending "want of power" in carrying out his project. Prudence, circumspection, humility and awe are the culprits which, the poet asserts, are

responsible for his mental stagnation is "blank reserve" and lead him astray on his journey towards "simplicity and self-presented truth." What Wordsworth overlooks, however, is the crucial fact that *the real impediment to the growth of his poetic mind lies in the inadequacy of the poetic mind with its concept.* The most important aspect of the Hegelian dialectics, according to Žižek, is the reversal whereby the failure of the particular to correspond to its universal, of the object to correspond to its concept, is reflected back into the universal concept itself. Apropos of explaining the Hegelian critique of Kant's notion of the sublime, Žižek (2008b) claims:

> *In Kant's view, the whole movement which brings forth the feeling of the Sublime concerns only our subjective reflection external to the Thing, not the Thing-in-itself – that is, it represents only the way we, as finite subjects caught in the limits of our phenomenal experience, can mark in a negative mode the dimension of the trans-phenomenal Thing. In Hegel, however, this movement is an immediate reflexive determination of the Thing-in-itself – that is, the Thing is nothing but this reflexive movement. (p. 242)*

That is to say, whereas Kant believes that the feeling of the sublime is evoked when the appearance, or phenomenon, fails to correspond to the Thing-in-itself, to the Thing which persists in its positivity beyond the veil of appearance, for Hegel, the Thing is nothing but this radical negativity, nothing but this inadequacy of the appearance to its concept. In other words, for Hegel, the Kantian Thing is nothing but the gap which forever separates the object from its concept, and the sublime is an object which gives body to and materializes this unbridgeable gap, an embodiment of the lack which is responsible for the ultimate failure of the symbolic order (Žižek, 2006, p. 25). From another perspective, the sublime object is the "negative" embodiment of our desire for unity and perfection. It is an object which gives us the promise that perfection is in itself possible, that one day we will be able to realize the concept, or the Idea, only if we manage to surmount the obstacles which for now impede our way to reach the ultimate object of our desire. The sublime object is our excuse for not being able, for having failed, to fulfill our dreams.

Wordsworth laments over his inability to realize the ideal poetic genius, over his failure to actualize the potentialities of his imaginative faculty. For him, it is

possible to surpass the gap between concept and its realization, between universal and particular. For this reason, he locates the source of his failure in external elements, the most important of which being his forced relocation from the "rural" Lake District to the city. City, in a sense, turns into a sublime object for the mature Wordsworth, who, looking back over the years, sees his poetic life as one of disappointment and defeat. If he had not been removed from the arms of Nature, had he not been delivered to the custody of the dull and tiresome life of city so early in his life, he would have reached the summit of poetry and would not have sunken into mental torpor and lethargy. City, therefore, turns into the nodal point, the *point de capiton*, which gives coherence and unity to Wordsworth's picture of life. Without it, without its quilting function, the world which he has made for himself falls to the ground. In other words, city is sublimated in order to cover the empty place of the Thing, in order to justify the inherent failure of Nature.

For Wordsworth, Nature is the epitome of Divine creation; it is the "breath of God," "His pure Word by miracle revealed." As such, there are no deficiencies in Nature, and whatever defects we find in it are due to *our* inability to have an insight into the overall divine framework and plan. In a celebrated passage of the first book of *The Prelude*, Wordsworth describes

his childhood experience of encountering the Sublime, of how one night he stole a boat and rowed across the Lake, when all of a sudden he found himself in the presence of a high peak beyond the lake, which seemed to menace him because of stealing the boat:

…lustily
I dipped my oars into the silent Lake,
And, as I rose upon the stroke, my Boat
Went heaving through the water, like a Swan;
When from behind that craggy Steep, till then
The bound of the horizon, a huge Cliff,
As if with voluntary power instinct,
Upreared its head. I struck, and struck again,
And, growing still in stature, the huge Cliff
Rose up between me and the stars, and still,
With measured motion, like a living thing,

Strode after me. With trembling hands I turned,
And through the silent water stole my way
Back to the Cavern of the Willow tree.
There, in her mooring-place, I left my Bark,
And, through the meadows homeward went, with grave
And serious thoughts. (Book I, ll. pp. 401-417)

This passage helps us analyze the concept of the Sublime under a new light, namely, its relation to the Lacanian notions of castration and oedipal complex. Of course, one cannot fail to see in this passage a dramatization of the oedipal scenario, whereby the imaginary mother-child unity is disturbed, once and for all, due to the intrusion of the paternal metaphor. The horrid and huge shape of the "peak" disturbs the peaceful nocturnal journey across the Lake. The young poet, dipping his oar "lustily" in the water, moving swan-like on the surface of the Lake under the influence of the moon and the stars, seems to be unconsciously acting out and enjoying the imaginary unity with the mother. However, all of a sudden, he witnesses the "grim shape" which towers up "between me and the stars," preventing him from getting closer to his destination which, read analytically, is the ultimate union with the mother. In other words, through its intrusion, the paternal metaphor of the mountain peak, with its erect and phallic shape, signifies the impossibility of reaching the ultimate object of desire, namely, the mother. The young Wordsworth, afraid at the sight of this unwelcome intruder, finds no way other than turning back to the shore and putting an abrupt end to his quest.

The effect of this encounter with the sublime image of the mountain, that is, the effect of castration on the psychical world of the poet is keenly described in the lines which follow. From the moment, up until many days, nothing seems to be the same as before. Wordsworth sees everything under a new light. The pleasant images of nature evacuate his mind, and nothing occupies their place but "huge and mighty forms" which remind him of the frightening shape of the mountain peak. In analytic terms, all the imaginary objects of the child's world are mortified by the intrusion of the Name of the Father, by the intrusion of signifiers occupying the place of imaginary objects. (Fink, 1995, p.120). The disintegration of the imaginary unity separates, forever, the child from ultimate union with the mother, putting her beyond reach as the impossible/prohibited object of desire. And it is precisely at this point

that the psychoanalytical concept of castration reveals its relevance to the argument of the sublime. In a sense, the sublime disturbs the relationship between the subject and Nature, creating a discordance at the heart of the supposedly harmonious connection between human beings and the surrounding world. That is to say, the experience of encountering the sublime introduces a gap into the subject's conception of Nature as a comprehensible totality. All of a sudden, the subject finds himself in the presence of an unknowable, incomprehensible phenomenon which does not fit into any existing system of signification. At this moment, Nature no longer functions as the passive object of perception, exposed before the subject's gaze; rather, it turns into a blot which blurs the subject's overall appreciation of the picture. (Lacan, 1998, p. 96). In Lacanian terms, the moment of the encounter with the sublime is the moment of Nature's 'de-subjectivization', the moment Nature turns the gaze back onto the subject himself. (Žižek, 1992, p. 186). The subject stares at the scene, yet its signification eludes him. It is at such a moment that the subject realizes the insufficiency of his mental apparatus to acquire a comprehensive knowledge of Nature, that he becomes aware of the radical impossibility which marks the core of the world around him. In other words, the effect of encountering the sublime is the same as the effect of castration, albeit on the optical and intellectual plane. The image of the sublime (which is almost exclusively of a phallic shape) creates a rupture in the subject's unity with the Mother Nature, turning the prospect of an ultimate possession of her into a mirage.

In yet another sense, the sublime is the moment of the intrusion of the real into the texture of the symbolic order, the moment when the subject recognizes the elemental nullity and the radical impotency of the Other, when it becomes apparent "that the Other (the symbolic order) is only a fiction" (Žižek, 2008a, p. 251). The sublime resembles the real in that, like the latter, it disturbs the symbolic reality by its obstinate resistance to interpretation, by its aversion to all the categories of understanding. It materializes the lack in Nature and gives body to the fact that Nature (as a consistent, totally comprehensible whole) does not exist. Thus, the ideal picture which the subject ever drew from Nature melts into thin air, and what remains is nothing other than the strange and morbid pile of objects which no longer signify anything. Harmony and order give their place to discord and chaos, leaving behind nothing but the fanciful illusion of a prelapsarian, primordial, time when

humanity was one with Nature, when human beings recognized themselves in the universe.

Therefore, the sublime can be approached from two different points. First, it can be conceived as the symbolic intrusion into the imaginary world of the subject, rifting the sense of unity and perfection, and relocating the object of desire from the imaginary to the symbolic order. Next, it can be perceived as the intrusion of the real into the symbolic order, as the lack at the heart of the Other, making a perfect correspondence of the signifying elements and the ultimate achievement of the object of desire impossible. What links these two different significations of the sublime together is the crucial realization that, in both cases, the sublime holds the place of illusory and non-existent states. In the first case, the sense of unity is an illusion, a state which has not existed in the first place. That is to say, the imaginary unity with the object of desire, later on disturbed by the intrusion of a third party, is a fantasy of a subject who has always found himself lacking and in search of the ultimate object of desire. In the second case, the real is also nothing but a subsequent creation of the symbolic order, the effect of the system of signification rather than its cause. In other words, the real comes into being the moment the symbolic order finds itself incapable of corresponding to its notion, the moment it recognizes the impossibility of perfect signification. In both cases, the sublime is a fantasy creation in order to justify the present state of things. The sublime did not put an end to an existing unity, neither did it create a hole at the heart of the symbolic system. It is, as Žižek often points out (2008b*,* p. 59), the case of a cause which comes after its effect: first we find the lack and only then do we look for its cause.

It is in this sense that we find the sublime experience appealing and appalling at the same time. Appalling because it disintegrates our illusory sense of unity with Nature; appealing because it gives us a chance to sink into our illusion, giving us a breathing space against the dark and bitter truth of our being. In other words, the sublime points to a state beyond itself, to a realm in which once we abided and we are now fallen from it. In a sense, it promises us that, behind the appearance of this world, filled as it is with wants and lacks, there is the Thing-in-itself, i.e. the ultimate object of our desire. The sublime, as it were, takes the full responsibility of our fallen, split, existence on itself, thus saving us from realizing the traumatic truth of the Other's ontological incompleteness. For Hegel, it was

Kant's attempt to avoid facing this traumatic truth that led to his supposition of the Thing-in-itself as a positive being forever excluded from our comprehension. The weight of acknowledging the intrinsic failure of reality was so heavy for Kant that he was more than ready to attribute the apparent inconsistencies of the system to his own limited capacity to understand the Thing-in-itself. The sublime, for Kant, marked the failure of our cognitive power to encompass the boundless realm of the real, to reach behind the phenomena to the noumenal realm. For Hegel, however, the sublime does not point to anything beyond itself. The supersensible Thing, Hegel argues, is nothing but a fantasy creation of philosophy in order to account for the present state of things. (Žižek, 2008b, p. 194). In ideological terms, the sublime is a privileged object which justifies the apparent failure of the social field, by giving body to the cause of our social problems, whereby, in an act of ideological inversion, the lack, inherent to the system, is projected onto an external object.

The effect of the encounter with the sublime, however, is passed unnoticed by the poet. His strong belief in the perfection of Nature prevents him from acknowledging the fact that perhaps the reason behind the failure of his poetic mind has not been anything *other than the failure of his source of inspiration itself*. Wordsworth's failure, he believes, is the effect of his separation from Nature and being overwhelmed in the trivialities of academic life and, later on, the life of the metropolitan London. The source of failure cannot be Nature since Nature is the mighty work of the Universal Wisdom, not the vulgar work of humanity. Therefore, phenomena such as the sublime, which seem at first glance to deviate from the immutable laws of Nature, turn out to be necessary elements in the education and the purification of the soul. The lines, preceding the passage of the encounter with the sublime, best portray Wordsworth's view on such apparently problematic experiences:

...there is a dark
Inscrutable workmanship that reconciles
Discordant elements, makes them cling together
In one society. How strange that all
The terrors, pains, and early miseries,
Regrets, vexations, lassitudes, interfused
Within my mind, should e'er have borne a part,

And that a needful part, in making up
The calm existence that is mine when I
Am worthy of myself! Praise to the end!
Thanks to the means which Nature deigned to employ! (Book I, ll.
pp. 352-366; emphasis added)

Of course, Wordsworth is right in his assertion that there is "a dark inscrutable workmanship" which is responsible for the reconciliation of the apparent conflicts, turning discordant sounds and voices into a harmonious melody. From a Marxist vantage, such a conciliatory and pacifying role is unequivocally played by ideology. It is through ideology and ideological fantasy that we learn to see the world as we do, that we learn to see peace in conflict, happiness in misery, and victory in defeat. (Žižek, 2008b, p. 142). Nature does not speak or act by itself; it is we, the subjects of ideology, who give voice to the otherwise mute mountains, rivers, and trees. In the Preface to the Second Edition of *Lyrical Ballads*, Wordsworth describes the principal object of the work to have been to select "incidents and situation from common life, and relate or describe them, throughout, as far as was possible in a selection of language really used by men, and, at the same time, to throw over them a certain colouring of imagination, *whereby ordinary things should be presented to the mind in an unusual aspect*" (Brett & Jones, 1991, p. 236). Wordsworth seems to be well aware that, without "a certain colouring of imagination," natural objects fail to rouse the human mind from torpor and lassitude, that Nature, in itself, is nothing but a "neutral" proximity of objects, incapable to carry out the lofty project of the universal guidance of mankind. Therefore, the "ghostly language of the ancient earth" is nothing other than the voice of the poet himself which is projected to the mute and silent Nature, proclaiming ideological "fictions" in the name of universal facts and "truths."

However, in spite of his consciousness and awareness, Wordsworth follows the logic of the fetishistic disavowal (Žižek, 2008b, p. 12), conveyed by the formula: "I know very well that [Nature is in itself incapable of communicating any truths to human beings;] however, I act as if [it is the ultimate guarantee of truth, the only medium which reveals divine wisdom and could educate and purify the human soul]." In other words, Wordsworth knows very well that Nature is not a unified and complete whole, that it is a lacking structure like all the other systems (his childhood

experience of encountering the sublime mountain peak at least points to this fact); however, in order to maintain the integrity of his psychical world, in order to save his life from collapsing into the abyss of psychotic disturbances, he *acts as if he does not know*. Here we are at the exact opposite pole of Marx's classical definition of ideology as "they do not know it, but they are doing it." As Žižek (2008b) puts it, the true formula of an ideological consciousness should put the stress not on "knowledge" but on the "act", that while people *know*, for instance, that money in itself is nothing but a worthless piece of paper, nonetheless they *act* as if it really possessed an intrinsic value of its own. (p. 27).

In the same vein, in order to give validity and a sense of consecration to his ideological inclinations, Wordsworth appeals to the universal wisdom of Nature and acts accordingly, in spite of his knowledge that there is no unique feature which would distinguish Nature from, say, City as such. In order to prove this point, it is enough to pay close attention to the passages in which Wordsworth addresses his life-long friend and companion, and the main addressee of *The Prelude*, Samuel Taylor Coleridge (1772-1834). Unlike Wordsworth who spent his childhood and adolescence in a rural area, among natural objects and scenes, Coleridge grew up in London, where no trace of Nature could ever be found. However, in spite of being thus deprived, Coleridge manages to rise to the same heights as those by Wordsworth:

Thou, my Friend! wert reared
In the great City, 'mid far other scenes;
But we, by different roads at length have gained
The self-same bourne. And for this cause to Thee
I speak, unapprehensive of contempt,
The insinuated scoff of coward tongues,
And all that silent language which so oft
In conversation betwixt man and man
Blots from the human countenance all trace
Of beauty and of love. For Thou hast sought
The truth in solitude… (Book III, ll. pp. 466-475)

In fact, these lines create an oppositional atmosphere within the entirety of the poem, since they reveal Wordsworth's belief in the possibility of gaining

"genuine" and "pure" knowledge far away from the natural domain, in the middle of the trivialities of the urban life. That is to say, it is not necessary to dwell constantly among natural objects, to have perpetual communication with birds, mountains, flowers, trees, rivers and lakes, in order to find truth and achieve greatness of mind and soul. Here, "solitude" takes Nature's place in leading the soul towards perfection. It is in solitude, Wordsworth claims, that Coleridge has found truth, in those moments when the soul shrinks back from the outside world and plunges deep into the remote recesses of the mind. Therefore, it becomes clear that Wordsworth's tirade against the city as the reason behind his mental sluggishness and torpidity cannot be justified.

If Coleridge could survive the hustle and bustle of city life and keep his imaginative faculty sharp and keen, then it turns out that City is nothing but a scapegoat which would account for Wordsworth's failure in fulfilling his dreams. In other words, on his way towards poetical and spiritual perfection, Wordsworth reaches an untraversable gap which impedes his progress, where, in an act of ideological inversion, he externalizes the sources of failure. For him, City turns into a sublime object (albeit of a negative nature), giving body to all the internal gaps and inconsistencies which have brought about his imaginative and poetic downfall. And this best clarifies the elevated place of City in the poet's psychical economy, since, the moment it would dissipate and cease to function as the cause of failure, Wordsworth's world would fall apart into pieces. That is, it makes it possible for Wordsworth to escape the traumatic knowledge that there is no way leading to perfection, that all the attempts to reach wholeness and integrity are doomed to failure.

Conclusion

Wordsworth never seemed to stop believing in the truth of the sublime objects, to stop searching for the roots of failure outside the ideological field: the lesson which he overlooked when he was a child, when he perceived the negative force inside Nature itself in the experience of the sublime shape of the mountain peak, and the one he ignored when he witnessed the failure of the ideals of the French Revolution. Robespierre, Napoleon, City, all were surrogates for the system's internal impossibility, ordinary object which by historical chance were

elected to materialize the Other's inherent infirmity and non-existence. Wordsworth's seeming ignorance of this crucial fact was the main reason behind his life-long mental and spiritual crises which showed themselves more and more as he grew more and more in age. From psychoanalytical perspective, Wordsworth failed in traversing his fundamental fantasy, in coming to terms with the ontological non-existence of the Other, with the primordial non-closure of the signifying system. Thus, we can consider *The Prelude* as Wordsworth's desperate attempt to save his ideological fantasy from falling to the ground, an attempt to prove both to himself and to the Other (materialized in the person of Coleridge, the absent addressee of the poem) that the Other exists, to prove that behind the disintegrated and lacking reality there is a unified and complete Whole, which is accessible through active inherence to and honest discipleship of Nature. *The Prelude* is the creation of an uneasy mind, of the mind of a poet whose (albeit illusory) belief in the perfection of Nature as the ultimate revelation of the Divine Spirit has helped him to rise above the ideological turbulences of his time, to find peace and solace though "[t]his Age fall back to idolatry, / Though Men return to servitude as fast / as the tide ebbs" (Book XIII, ll. pp. 432-434). By the end of the poem, Wordsworth has turned into the "Prophet of Nature," and *The Prelude* the guiding principle of humanity, "A lasting inspiration, sanctified / by reason, blest by faith":

> *what we have loved*
> *Others will love; and we may teach them how,*
> *Instruct them how the mind of man becomes*
> *A thousand times more beautiful than the earth*
> *On which he dwells, above this Frame of things*
> *(Which, 'mid all revolutions in the hopes*
> *And fears of men, doth still remain unchanged)*
> *In beauty exalted, as it is itself*
> *Of substance and of fabric more divine. (Book XIII, ll. pp. 444-452)*

References

Abrams, M. H. (1973). *Natural supernaturalism: Tradition and revolution in romantic literature*. Norton.

Brett, R. L., & Jones, A. R. (1991). *Lyrical ballads: William Wordsworth and Samuel Taylor Coleridge* (2nd ed.). Routledge.

Day, A. (2001). *Romanticism*. Routledge.

Fink, B. (1995). *The Lacanian subject: Between language and jouissance*. Princeton University Press.

Fink, B. (2007). *Fundamentals of psychoanalytic technique.* W. W. Norton & Company.

Hartman, G. (1964). *Wordsworth's poetry 1787-1814.* Yale University Press.

Heffernan. J. A. W. (1967). Wordsworth on the sublime: The quest for interfusion. *Studies in English Literature, 1500-1900, 7*(4), 605-615. https://doi.org/10.2307/449528

Hopkins, B. (1994). Wordsworth's voices: Ideology and self-critique in "The Prelude." *Studies in Romanticism*, *33*(2), 279-299. https://doi.org/10.2307/25601060

Kelley, T. M. (1984) Wordsworth, Kant, and the romantic sublime. *Philological Quarterly*, *63*(1), 130.

Lacan, J. (1998). *The seminar of Jacques Lacan: The four fundamental concepts of psychoanalysis (Book XI)* (Revised ed.). W. W. Norton & Company.

Myers, B. P. (2008). Wordsworth's financial sublime: Money and meaning in book VII of "The Prelude." *South Central Review, 25*(2), 80-90. http://www.jstor.org/stable/40211265

Owen, W. (1973). The sublime and the beautiful in *The Prelude*. *The Wordsworth Circle, 4*(2), 67-86. http://www.jstor.org/stable/24039116

Stang, R. (1966). The false dawn: A study of the opening of Wordsworth's *The Prelude*. *ELH*, *33*(1), 53-65. https://doi.org/10.2307/2872134

Stoddard, E. (1985). Flashes of the invisible world: Reading *The Prelude* in the context of the Kantian sublime. *The Wordsworth Circle, 16*(1), 32-37. https://www.jstor.org/stable/24040569

Weiskel, T. (1976) *The romantic sublime: Studies in the structure and psychology of transcendence.* Johns Hopkins University Press.

Wlecke, Albert O. (1973). *Wordsworth and the Sublime*. University of California.

Wordsworth, W. (1995) *The Prelude*. London: Penguin.

Žižek, S. (1992). *Enjoy your symptom*. Routledge.

Žižek, S. (2006). *The parallax view*. MIT Press.

Žižek, S. (2008a). *For they know not what they do*. Verso.

Žižek, S. (2008b). *The sublime object of ideology* (2nd ed.). Verso.

Journal of Language Horizons

Scientefic Quarterly Journal of Language Horizons, *Alzahra University*
Volume 7, Issue 1, Spring 2023 (Serial No. 15)

Proactive Form-Focused Instruction in a Flipped Classroom: Implicit and Explicit Grammar Knowledge Development and Retention

Research Article
pp. 107-134

Ahmad Khalifeh[1]

Mohammad Bavali[2*]

Ehsan Rassaei[3]

Received: 2021/10/04 Accepted: 2022/02/12

Abstract

The present study investigated the possible effects of proactive form-focused instruction on Iranian EFL learners' explicit and implicit grammar knowledge improvement in a flipped and a traditional class and also examined the effect of flipped class on explicit and implicit knowledge retention. Two classes were randomly assigned to an experimental (n = 31) and a control (n = 28) group. The former received proactive form-focused instruction through metalinguistic explanation in a flipped class, whereas the latter attended a traditional grammar class. Two implicit and two explicit knowledge tests were used to evaluate the learners' explicit and implicit grammar knowledge development and retention. A set of pretests and two sets of posttests were administered immediately and four weeks after the last treatment session. Two mixed 3x2 multifactorial ANOVAs and *post hoc* tests were run to spot the differences in the measures between and within the groups. The results showed that both classes improved significantly in terms of learners' implicit and explicit grammar knowledge.

[*] Corresponding Author
[1] Ph.D. Candidate, Department of English, Faculty of Humanities and Social Sciences, Shiraz Branch, Islamic Azad University, Shiraz, Iran. ahmadkhalifeh@iaubushehr.ac.ir
[2]Assistant Professor, Department of English, Faculty of Humanities and Social Sciences, Shiraz Branch, Islamic Azad University, Shiraz, Iran. mbvl57@gmail.com
[3] Associate Professor, English Department, Faculty of Humanities and Social Sciences, Shiraz Branch, Islamic Azad University, Shiraz, Iran. ehsan.rassaei@majancollege.edu.om

DOI: 10.22051/lghor.2022.37984.1571
DOR: 20.1001.1.2588350.2023.7.1.5.3

Furthermore, the experimental group showed a higher rate of knowledge retention. The findings can help and ensure language program designers, educational policymakers, and language teachers to capitalize on flipped courses to teach grammar effectively and help language learners retain both explicit and implicit grammar knowledge longer.

Keywords: explicit, flipped, form-focused, implicit, retention

Introduction

Numerous studies in the literature have investigated Form-Focused Instruction (FFI) and the role it may play in Second Language (L2) development (e.g., Khezrlou, 2021; Toni & Hassaskhah, 2018; Trahey & Spada, 2020). The general theme that has emerged indicates it can affect L2 acquisition positively, as supported by several studies and meta-analyses (e.g., Kang et al., 2019; Saito, 2013; Spada, 2011). However, most previous studies have been conducted in traditional face-to-face classes (e.g., Lee, 2021; Lindseth, 2016; Xu & Li, 2021), and little information exists about the effectiveness of FFI in technology-enhanced language learning environments, the importance of which is becoming increasingly apparent during the COVID-19 pandemic lockdown. Particularly, the literature is characterized by a dearth of studies focusing on FFI in technology-enhanced learning environments in an Iranian context where despite the widespread use of digital tools and social media, the integration of technological equipment in the educational contexts is unsatisfactory yet (Badrkhani, 2021, Rassaei, 2017). This weakness has been referred to in Rassaei (2017) when he says "... in Iran where despite mounting interest in CALL, the incorporation of digital and technological tools in educational contexts is still in its Infancy" (p. 2).

One of the educational environments supported by technology is the flipped class that has widely been studied in recent years. Studies on it, especially in Iran, have mainly focused on its possible effects on language proficiency, students' perceptions, opinions, motivation, engagement, critical thinking, self-efficacy, and autonomy, teachers' perceptions and views as well as comparison between a traditional class and a flipped class (e.g., Afzali & Izadpanah, 2021; Amiryousefi, 2019; Fathi & Rahimi, 2020; Izadpanah & Afzali, 2021; Lee & Wallace, 2018; Nourinezhad et al., 2022).

This new pedagogical approach can increase the input frequency by

engaging students in various educational activities in which the target structures can be incorporated. Research has also shown that technology-enhanced learning environments can increase learners' motivation as well as engagement with content that, in turn, can facilitate deeper understanding, improve memory performance, and enhance knowledge retention (Bouwmeester et al., 2019; Francis, 2017; Semb & Ellis, 1994; Woolfolk & Margetts, 2016). Because a flipped class seems to have the potential to provide these benefits, it is expected that knowledge acquired in a flipped class will be retained longer. However, in the literature, there are not many studies that have documented the direct effect of technology-enhanced learning environments on knowledge retention. This is particularly more evident in research on flipped classes.

Additionally, the effect of explicit instruction on enhancing implicit and explicit knowledge of an L2 has been extensively studied recently (e.g., Akakura, 2012; De Graaf, 1997; Ellis, 2004; Ellis, 2017; Li et al., 2019; Xu & Lyster, 2014). Nevertheless, the extent to which explicit instruction may yield explicit and/or implicit knowledge is still a controversial issue. There is a complex relationship between implicit/explicit instruction, learning, and knowledge. The relationship between these concepts is not as direct and straightforward as some expect and merits further research. Explicit language learning, which often results in metalinguistic rules, "is necessarily a conscious process and is generally intentional" (Ellis, 2009, p. 7), whereas implicit learning occurs without awareness when learners are immersed in a considerable amount of input (Ellis, 2009). Thus, it seems that implicit knowledge acquisition need more time (DeKeyser, 2003) and requires a larger amount of input. In addition, several researchers have argued that explicit learning might result from attention to L2 forms, negative evidence, and explicit instruction (e.g., Lichtman, 2016), whereas some others believe that consciously guided practice is one of the pedagogical strategies which can lead to unconscious implicit knowledge (e.g., Ellis, 2008). Therefore, so far, the results of investigations seem to be inconclusive. By increasing the intensity of instruction, exposing students to more input, and using a range of instructional activities inside and outside the classroom, flipped class is likely to have a contributing role in promoting students' both implicit and explicit knowledge that seems to be the primary goal of L2 instruction. Most, if not all, research on the effect of flipped class on knowledge

retention has been conducted in medicine (e.g., Bouwmeester et al., 2019; Graham et al., 2019) and it seems that the language teaching field suffers from a lack of research in this area. Thus, the present study has a pioneering role in investigating the effect of a flipped class on the retention of English grammar knowledge. In addition, as mentioned above, most research on form-focused and explicit instruction have been conducted in traditional classes.

Given the foregoing issues, the present study was intended to examine and compare the possible effects of explicit instruction on students' implicit and explicit knowledge of grammar in both flipped and traditional courses in an Iranian higher education context. In addition, it sought to gain further understanding of the flipped class effect on English grammar knowledge retention.

Literature Review

Flipped Classroom

Nowadays, technology has turned into an integral part of language learning and teaching environments, changing the ways instruction is fulfilled. One of the fairly new educational innovations is the flipped classroom, which as a type of blended learning, can be employed to provide learners with various types of instruction. The concept of a flipped class is simply defined as "that which is traditionally done in class is now done at home, and that which is traditionally done as homework is now completed in class" (Bergmann & Sams, 2012, p. 13). Thus, it reverses or "flips" the regular method of teaching by delivering instruction to students at home through prerecorded video lectures, podcasts, slides, and texts and moves the homework phase into the classroom in which students engage in assignments, problem-solving activities, discussions, games, analyses, mastery quizzes, and face-to-face interaction with peers and teachers (Roehling, 2018), whereas the teacher "instead of being the 'sage on the stage,' functions as a 'guide on the side,' facilitating learning in less directive ways" (King, 1993, p. 30).

A brief literature review shows that the flipped class has mainly been used in fields such as mathematics, engineering, statistics, technology, science, and medicine (e.g., Clark, 2015). In addition, it has recently been adopted in the fields of teacher education and language teaching (e.g., Turan & Akdag-Cimen, 2020) and widely utilized for teaching different language components and skills (Kang, 2015;

Webb & Doman, 2016). The general theme on the use of flipped instruction in different fields suggests that it can result in improved academic performance, positive students' perception, increased engagement, enhanced motivation, and autonomy (e.g., Cilliers & Pylman, 2020; O'Flaherty & Phillips, 2015; Zainuddin & Perera, 2019). Therefore, it appears to be a potential pedagogical strategy for teaching a language. Furthermore, previous investigations show that only a few studies, mainly in the fields of medicine and mathematics, have examined its possible effect on reducing the rate of memory loss and enhancing knowledge retention (e.g., Bouwmeester et al., 2019; Morton & Colbert-Getz, 2016) and the effect of a flipped class on linguistic knowledge retention has not been studied yet in the field of language teaching. However, the possible effects of this type of class on learning English grammar in EFL contexts have been examined in a number of papers, some of which will be cited in *Flipped Classroom and FFI*.

Explicit and Implicit Knowledge and Instruction

Another major theme in SLA research is the explicit-implicit instruction dichotomy. Explicit instruction involves rule explanation (Housen & Pierrard, 2005) and encouraging learners to develop "metalinguistic awareness of the rule" (Ellis, 2009, p.17). Implicit instruction, however, is intended to help learners deduce grammar rules without awareness by providing them with sufficient exemplars of a rule (Ellis, 2009). These two types of instruction are further categorized into reactive and proactive. The former occurs mostly as corrective feedback to focus learners' attention on the target form during interaction (Lyster, 2007), whereas the latter involves "preplanned instruction designed to enable students to notice and to use target language features that might otherwise not be used or even noticed in classroom discourse" (Lyster, 2007, p. 44). Furthermore, "explicit L2 knowledge is the declarative and often anomalous knowledge of the phonological, lexical, grammatical, pragmatic, and sociocritical features of an L2 together with the metalanguage for labeling this knowledge. It is held consciously and is learnable and verbalizable" (Ellis, 2004, pp. 244-5), but implicit language knowledge, which cannot be normally verbalized, is internalized subconscious knowledge tapped into by real-time or spontaneous language tasks (Ellis, 2009).

A review of previous research indicates that implicit instruction generally

seems to be less effective than explicit instruction in helping learners to acquire L2 forms (e.g., Akakura, 2012; De Graaff, 1997; Kang, et al., 2019; Norris & Ortega, 2000). A number of studies have also shown that explicit knowledge of grammar can assist language learners in noticing a target form and gradually acquire it so that it changes into implicit knowledge used for spontaneous production (Ranta & Lyster, 2018). Akakura (2012), for example, investigated how explicit instruction can affect L2 students' implicit and explicit knowledge of English articles and found enduring effects for explicit instruction on explicit and implicit knowledge. In another study, De Graaff (1997) examined the effect of explicit instruction, complexity, and morphology/syntax on the acquisition of four L2 structures. His findings showed that explicit instruction promotes L2 acquisition. In a more recent study, Xu and Lyster (2014) focused on the possible differential effect of explicit FFI on using morphosyntactic forms in L2 oral production. They reported that FFI improved the target form use and its effect was moderated by the regularity and complexity of the morphosyntactic forms.

On the other hand, several studies have questioned the facilitative role of explicit instruction in L2 acquisition. Li et al. (2019), for example, studied the effect of pre-task explicit teaching on a focused task and indicated that it resulted in negative global effects on the learners' L2 production. Likewise, Sanz and Morgan-Short (2004) studied the potential effect of computer-assisted explicit instruction on learning Spanish word order and concluded that explicit instruction may not necessarily promote L2 acquisition. Previous studies also indicate the effectiveness of explicit instruction depends upon several factors like the target structure complexity, its availability in non-instructional input, and instruction intensity (Ellis, 2002; Hulstijn & de Graaff, 1994). A flipped classroom can provide students with more time and opportunity, in the form of activities carried out inside and outside the classroom, to learn the intended topics. Thus, it appears to have enough potential to increase the amount and probably the effectiveness of explicit instruction, compared to a traditional class.

Flipped Classroom and FFI

Whereas most previous studies of FFI were carried out in traditional L2 classes, many studies have looked into the effects of various forms of FFI in

technology-enhanced courses (e.g., Rassaei, 2017). However, compared to other forms of technology-enhanced classes little experimental research has been conducted to study the effects of FFI in a flipped class. Asaka et al. (2018) designed a study in which a flipped class was employed to teach English present progressive and regular verb past tenses to a group of Japanese high-school students. The results showed that the flipped class had improved the students' grammar knowledge and speaking ability. However, when the outcome was compared with that of a control group that had received traditional instruction the difference was not statistically significant. Noroozi et al. (2020) also studied the effects of a flipped class on Iranian EFL learners' acquisition of English conditionals. Their findings implied that explicit instruction in the flipped class was more helpful than explicit instruction in a regular class on both immediate and delayed posttests. In addition, Bezzazi (2019) examined the effect of a flipped class on EFL learners' grammar knowledge development. The present, past, present perfect tenses, conditionals types 1 and 2, and the passive voice were taught and tested over 10 weeks. The results indicated that the flipped class was more effective than the regular instruction regarding teaching English grammar. Kang (2015) also compared two general English courses in a regular and a flipped classroom. Pre/Post-tests were used to compare the students' grammar and vocabulary knowledge before and after the instruction. The findings suggested that only flipped instruction resulted in statistically significant changes in both lexical and grammatical knowledge. Similarly, Webb and Doman (2016) studied the effect of a flipped class on the students' grammar knowledge. Their results showed that students in the flipped group outperformed their counterparts in the control group. Bulut and Kocoglu (2020) also evaluated the effect of a flipped class on EFL students' grammar knowledge and concluded that the flipped group's mean score was higher than the control group's and the difference between them was statistically significant. In another study conducted in an Iranian EFL context, Izadpanah & Afzali (2021) investigated the possible effect of a flipped class on critical thinking and grammar knowledge development. The flipped class was employed to teach the experimental group whereas the control group received instruction using the traditional way of teaching grammar. At the end of the study the flipped class proved to be more effective than the other one.

In contrast, Liu et al. (2019) used a student response system in a flipped

class to teach English grammar. They found that students' self-efficacy, class participation, learning motivation, and engagement in activities improved in the flipped classroom, whereas it proved ineffective in increasing students' grammar knowledge. As the literature suggests studies devoted to the flipped class effect on grammar knowledge development have resulted in inconclusive and mixed results. Furthermore, the grammar knowledge reported in these studies has generally been treated as a unitary construct and its quantity depends heavily on the tests used to measure it. To deal with the gaps mentioned above, the present study seeks to answer the following research questions:

1. Is there a statistically significant difference between the performance of students who received FFI in the flipped class and that of students who were instructed in the traditional class on explicit and implicit grammar knowledge posttests?
2. Is there a statistically significant difference between the performance of students who received FFI in the flipped class and that of students who were instructed in the traditional class on explicit and implicit grammar knowledge delayed posttests?

Method

Participants

Two intact classes including a total of 59 university freshmen studying English as a foreign language at two branches of Azad University in Iran participated in the study. They were selected on the basis of the convenience sampling method. All the participants were Iranian students with Farsi as their first language. Out of the 59 students, 41 were female and 18 were male undergraduates whose ages ranged from 19 to 31(M = 22). Both classes, which were taught by the same instructor, had to take a compulsory grammar course during the second semester of the 2019-2020 academic year. Only freshmen at the pre-intermediate level were selected in an effort to ensure that they had not yet acquired the target structures. They also completed consent forms and hereby declared their agreement to participate in the study.

Instruments

Two implicit knowledge tests (a Timed Grammaticality Judgement Test

(TGJT) and an Elicited Imitation Test (EIT)), two explicit knowledge tests (an Untimed Grammaticality Judgement Test (UGJT) and a Metalinguistic Knowledge Test (MKT)) and the Oxford Quick Placement Test (OQPT) were employed for the purposes of the study. Several previous studies have offered evidence on the construct validity of the explicit and implicit measures (e.g., Bowles, 2011; Ellis, 2009; Ellis & Loewen, 2007). In addition, four experienced English language teachers evaluated the content of the tests to ensure their content validity. Then, modifications were made in the test items to take care of their opinions.

OQPT. The OQPT, as a standard test that enjoys an acceptable level of validity and reliability, was used to determine the students' language proficiency level and help the researchers to select the participants. This is a paper-based version of the electronic Oxford Placement Test for learners of English. The OQPT consisted of 40 items. Its reliability measured by Cronbach's Alpha was $\alpha = 0.79$.

TGJT. The TGJT used as an implicit knowledge test included 25 items distributed randomly and displayed using PowerPoint slides. Of the 25 items, 13 statements were ungrammatical and 12 were grammatical. The amount of time allowed for the sentences varied between 6 and 9 seconds depending on each sentence length. Each correct item received 2 points while an incorrect or unanswered item was given a zero. An example of the sentences used in the TGJT is given below:

If I had seen him, I would have told you. *[Used in the delayed posttest]*

UGJT. The UGJT was constructed to tap the students' explicit knowledge of the target structures. This pen-and-paper test was identical to the TGJT in terms of the number of questions and grammar points examined. However, the number of ungrammatical and grammatical sentences was different. Of the 25 items, 11 were grammatical and 14 were ungrammatical. The participants were given an answer sheet including the instructions as well as the statements and were asked to judge them in terms of grammaticality without any time limits. Correct answers were given 2 points and incorrect ones 0 point.

MKT. The MKT designed to operationalize the construct of explicit knowledge comprised 25 items, too. Each item included two sentences, one of which was underlined and involved a grammatical error in using the target structures. The test-takers were asked to read the items, write down the grammatical

rules being violated in either English or Farsi, and correct the error. There was no time limit to answer. Students were given one point for stating the rule and one point for correcting the error. However, no score was assigned to only correcting the errors but not describing the rules. The following is a sample of sentences used in the MKT test.

We're out of the sugar. Put some cream in your coffee. *[Used in the immediate posttest]*

EIT. Finally, the EIT consisted of 25 belief statements involving 15 ungrammatical and 10 grammatical sentences. It was administered during one-on-one meetings between the researchers and the students who were asked to listen to 25 statements one at a time on a cellphone, indicate their opinion on an answer sheet which included "*Agree"*, "*Disagree"*, and "*No Idea"* options for each statement, and immediately repeat the statements in correct English. Each ungrammatical statement included only one error and each correct repetition was awarded two points. All the students' responses were audio-recorded for later analysis. A sample of the statements is given below.

Millions of people killed by cancer last year. *[Used in the pretest]*

Three parallel versions of each explicit and implicit test were developed to be used as pretests, immediate and delayed post-tests. The correlations between the control group's scores in the pretest and posttests were estimated

Table 1

Test-Retest Reliability Coefficients

Test	Pre - $Post_1$	Pre- $Post_2$	$Post_1$ - $Post_2$
TGJT	0.91	0.88	0.89
UGJT	0.77	0.87	0.82
MKT	0.84	0.81	0.96
EIT	0.86	0.88	0.97

Note. Pre = Pretest, $Post_1$= Immediate post-test, $Post_2$= Delayed post-test

for each implicit and explicit test to determine their test-retest reliability coefficients. Table 1 presents reliability coefficients for each test.

Data Collection Procedure

Pretreatment Stage

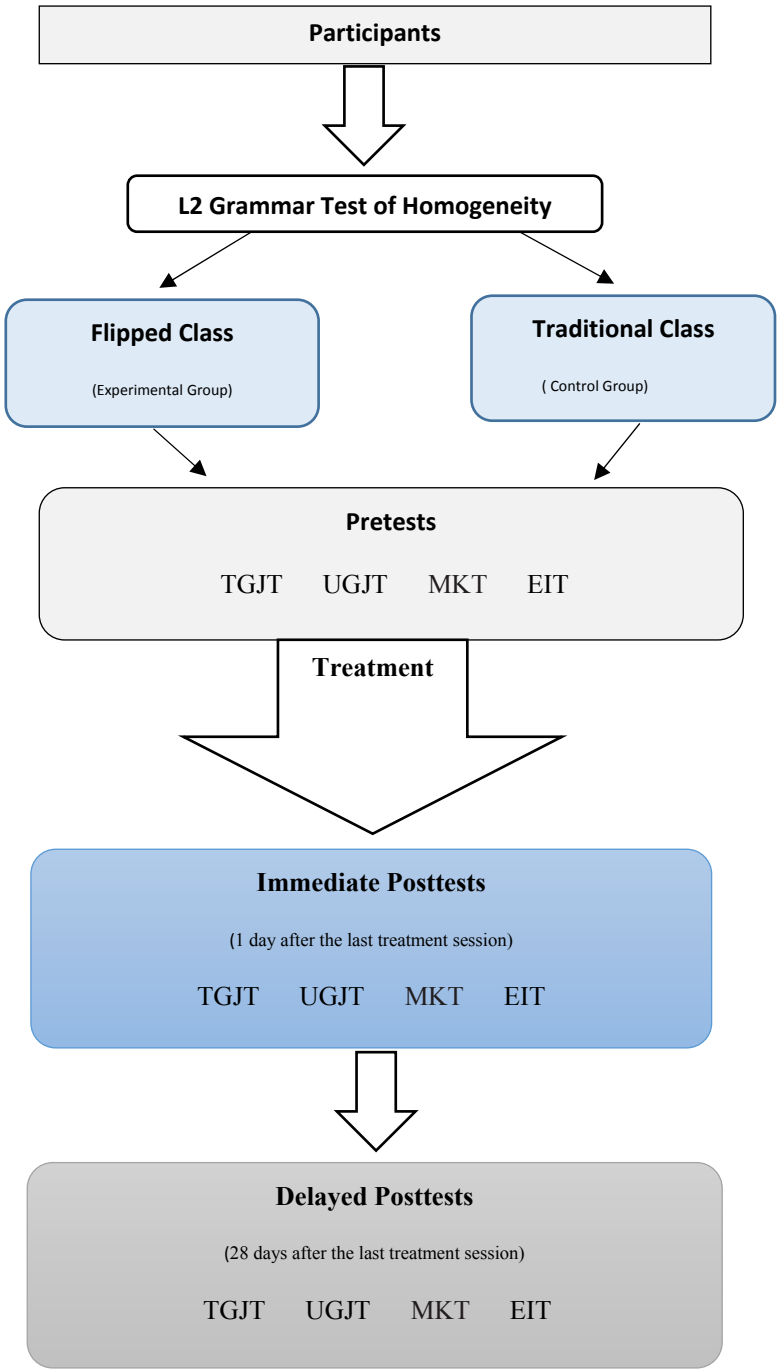
In order to gather the necessary data, the following procedure was followed. First, a group of 64 undergraduate TEFL students who were members of two different classes took the OQPT administered one week before beginning the instruction. Fifty-one students who scored between 18 and 30 (M = 23.84 and SD = 5.5) were selected as the participants at the pre-intermediate level. Then, the two classes were assigned randomly to a control (n = 28) and a treatment group (n = 31). Four days before starting the instruction, both groups took four pretests over three successive days. Afterward, both received instruction on the same grammar topics but in different ways for 11 weeks. The classes were held twice a week for both groups and every session lasted about 105 minutes, with a 10-minute break in between. The textbook (i.e., *English Grammar in Use*), instructor (i.e., one of the researchers), number of sessions (n = 22), and language of instruction (i.e., a mixture of Persian and English) were the same for both groups. A detailed description of the treatment and the control conditions follows. In addition, Figure 1 vividly depicts the data collection procedure followed in the present study.

Treatment Condition

Out-of-Class Phase. The researchers selected suitable videos through an extensive review of language learning websites. The main selection criteria were language, clarity, length, quality, pacing, and tone. In order to support the out-of-class part of the curriculum, the course content was delivered via *WhatsApp* to the students. Three or two days before holding each class the materials were sent to a WhatsApp group created for the purposes of the study. First, the videos and supplementary files including webpages, texts, and PowerPoint slides were uploaded one after another. Then, the day before the following session, a set of fill-in-the-blank or multiple-choice questions on the newly taught grammar structures were sent to the group. Each student was required to answer them and send back the answers to the instructor. They were informed that completing the tests constituted a portion of their total score.

Figure 1

Data Collection Procedure



In-Class Phase. Each class normally began with a brief review of the grammar points already presented by videos to clear up the students' likely problems and clarify any misunderstood points. Sometimes they had to take a short impromptu quiz on the target grammar points at the beginning of each class to ensure that all the students had watched the videos and studied the materials.

Then, they were asked to do the textbook exercises in pairs and volunteers took a turn reading out the completed exercises. In case of any problems, the instructor, in collaboration with other students, tried to remind them of the relevant grammar point. Afterward, the students, in pairs, would write English sentences using the new grammar structures, read them out to the class, and receive verbal feedback from peers and the instructor. Then, they were divided into groups of four and shown a series of pictures that depicted an event or simple story. Each group was required to write a brief story about them. The pictures and instructions were chosen in such a way that they could elicit the desired structures. After that, each group was given a chance to present their story to the class and get feedback from their classmates and the instructor who would often write their errors on the board and corrected them collaboratively. It should also be noted that some grammar games would replace the picture-cued storytelling task every other session. While they were completing the tasks, the instructor encouraged them to help and correct each other within the groups. As the final step, the group members would read their sentences aloud and get feedback. Although about ten grammar topics were presented during the semester, the study only focused on the passive voice, conditionals, and articles for practical problems and measurement issues.

Control Condition

The learners of the control condition attended a lecture-based grammar class twice a week, too. Each class began with a brief review of the key points presented during the previous session and proceeded with completing and checking the textbook exercises. Then, the instructor delivered a lecture during which new grammar points were presented explicitly in both Farsi and English. Afterward, a number of examples and further explanation were presented to the students. Next, he asked them to write a few sentences using the new structure. After that, students volunteered to read their sentences and received feedback from their classmates and

the instructor. Later on, a number of the book exercises were completed and the rest were assigned as homework.

Posttreatment Stage

Both groups took the immediate and delayed posttests a day and 28 days after the last class session, respectively. A review of the literature revealed that the time interval between delayed retention tests and the instruction is suggested to be two weeks or more (Haynie, 2007). The order of administering the tests was altered in the pretests and posttests to minimize the order effect.

Data Analysis

First, the students' scores on the TGJT and EIT were added together and an implicit knowledge total score was calculated. Similarly, their scores on the UGJT and MKT were added up and an explicit knowledge total score was computed for each participant, too. It should be noted that all reported total scores on pretests and posttests were out of 100. Then, to show that both groups were homogeneous regarding their explicit and implicit grammar knowledge before treatment, the participants' raw scores in the pretests were obtained, descriptive statistics were calculated for both groups, and two independent samples *t*-tests were run. Next, two mixed 3x2 multi-factorial ANOVAs and *post hoc* tests were carried out to answer the research questions. *SPSS 21* and *GPower 3.1* were employed to perform the analyses.

Results

Table 2 that provides descriptive statistics for learners' pretest, posttest, and delayed posttest scores in the explicit and implicit knowledge tests shows scores in both groups increased from the pretests to the posttests but deteriorated from the posttests to the delayed posttests. Previous research (e.g., Ebbinghaus, 1964) has shown that such reduction seems to be a natural phenomenon that occurs over time.

In addition, Figures 2 and 3 display changes in the groups' mean scores of implicit and explicit knowledge tests. Changes in scores from the pretests to the immediate posttests and from the immediate posttests to the delayed posttests represent learners' knowledge acquisition and knowledge retention, respectively.

The line graphs show that students in the flipped class obtained higher scores on both explicit and implicit tests. However, two independent samples *t*-tests on pretest scores showed no significant difference between the groups in both explicit, $t(47.73) = -.38$, $p = .70$ and implicit knowledge test scores, $t(57) = -1.13$, $p = .26$. This suggests that the two groups were homogenous before initiating the treatment.

Furthermore, two mixed 3x2 multi-factorial ANOVAs were carried out. Results showed that on the immediate posttests no statistically significant difference was found between the two teaching methods in improving the learners' explicit and implicit grammar knowledge, F_{exp} (1, 57) = 3.17, $p = .08$ and $F_{imp}(1, 57) = .88$, $p = .35$. However, the results indicated that there was a significant interaction effect between time and teaching method. This implied that students' explicit and implicit grammar knowledge improved differentially over time as a result of employing different teaching methods, $F_{exp}(1.58, 90.20) = 21.46$, $p < .001$, $\eta^2 = .27$ and $F_{imp}(1.58, 90.32) = 29.77$, $p < .001$, $\eta^2 = .34$.

Table 2

Descriptive Statistics of Explicit and Implicit Knowledge Test Scores

Knowledge			Pretest		Posttest 1		Posttest 2	
	Group	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Explicit Knowledge	Flipped	31	41.07	8.85	69.45	8.46	59.42	9.10
Scores	traditional	28	42.17	12.66	62.98	12.37	50.12	14.55
Implicit Knowledge	Flipped	31	32.93	11.70	61.66	10.30	55.48	10.77
Scores	traditional	28	36.76	14.26	57.39	12.90	47.13	14.75

Independent t-tests, as *post hoc* tests, revealed that the flipped and the traditional groups only significantly differed in explicit and implicit delayed posttests, $t_{exp}(44.45) = 2.90$, $p < .016$ and $t_{imp}(57) = 2.50$, $p < .016$. In addition, the within-group main effect of time was also significant in both explicit and implicit tests, $F_{exp}(1.58, 90.20) = 450.33$, $p < .001$, $\eta^2 = .88$ and $F_{imp}(1.58, 90.32) = 488.72$, $p < .001$, $\eta^2 = .89$. Two repeated-measures one-way ANOVAs were run to examine the explicit knowledge scores across the three test conditions for both flipped and traditional groups.

Figure 2

Explicit Knowledge Acquisition and Retention

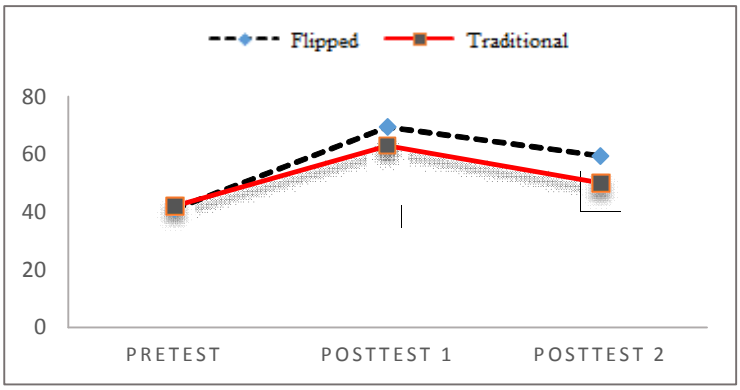
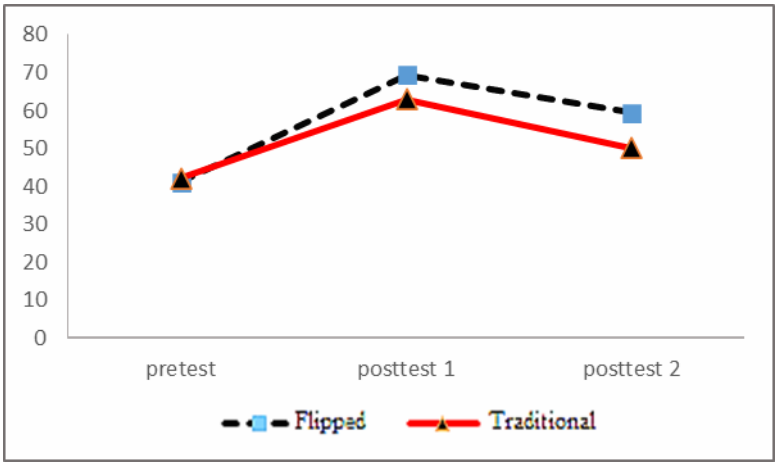


Figure 3

Implicit Knowledge Acquisition and Retention



Results showed a significant difference existed in explicit knowledge test scores across time for both flipped, F (1.43, 48.03) = 288.14, p < .001 and traditional groups, F(2, 54) = 181.32, p < .001. Also, two other repeated-measures one-way ANOVAs were carried out to investigate the implicit knowledge scores across the three test conditions for both groups. A significant difference was found in implicit knowledge test scores across time for both flipped, F(1.67, 50.37) = 413.17, p < .001 and traditional groups, F(1.41, 38.21) = 141.32, p < .001. Bonferroni *post hoc* tests showed significant differences in explicit knowledge scores between T1 - T2 (p <

.001), T1 - T3 ($p < .001$), and T2 - T3 ($p < .001$) for both flipped and traditional groups, where T1, T2, and T3 stand for the pretest, posttest, as well as delayed posttest, respectively. The second set of Bonferroni *post hoc* tests also demonstrated that there existed significant differences in implicit knowledge scores between T1 - T2 ($p < .001$), T1 - T3 ($p < .001$), and T2 - T3 ($p < .001$) for both flipped and traditional groups. Thus, the results indicate that both methods were effective in increasing the students' explicit and implicit knowledge.

Furthermore, to increase our understanding of the flipped class effect on knowledge retention, the ratio of the learners' delayed posttest scores over their immediate posttest scores was calculated for both implicit and explicit knowledge tests and the results were represented as percentage scores. Table 3 shows that the mean scores of students in the flipped class in both explicit as well as implicit knowledge retention tests are higher. The students in the flipped class retained 85% of their explicit grammar knowledge at the time of the delayed posttest, whereas those in the traditional group retained only 78%.

Regarding implicit knowledge retention, the numbers increased in both groups. While the learners in the flipped class succeeded in retaining 90% of their implicit grammar knowledge four weeks after the immediate posttest, their counterparts in the traditional class managed to preserve 80%. In general, the data indicate the superiority of the students of the flipped group in retaining both types of knowledge. Furthermore, implicit knowledge showed a higher retention rate compared to explicit knowledge in both classes.

Table 3

Proportion of Retained Explicit and Implicit Knowledge

Knowledge Type	Group	*n*	*M*	*SD*	*SEM*
Explicit Knowledge	Flipped	31	.85	.06	.011
	traditional	28	.78	.09	.017
Implicit Knowledge	Flipped	31	.90	.089	.016
	traditional	28	.80	.092	.017

Note. SEM= Standard Error of the Mean

Discussion

Findings revealed that explicit instruction in the flipped classroom was as effective as explicit instruction in the traditional class and effectiveness was not modulated by the teaching method although the bulk of the literature has reported the superiority of the flipped instruction in terms of improving student performance (e.g., Bezzazi, 2019; Bulut & Kocoglu, 2020; Fathi & Rahimi, 2020; Izadpanah & Afzali, 2021; Vaezi et al., 2019; Vitta & Al-Hoorie, 2020; Webb & Doman, 2016). What this study found is in agreement with the findings of a number of other studies in the language teaching field. Asaka et al. (2018), for example, found no significant difference between the students' grammar achievement in a flipped class and a traditional class. None of these studies, however, distinguish between different types of grammar knowledge.

There seem to be several possible explanations for the result obtained in the present study. First, the inconsistency found may pertain to different ways of executing the flipped class model. Although the general theoretical definitions of the concept presented in the literature appear to be broadly similar, it has been implemented in different ways in practice (e.g., McLaughlin, 2018). Second, as numerous studies (e.g. Lee & Wallace, 2018) have demonstrated flipped instruction probably needs time to reveal its beneficial effects. Therefore, it appears not to have a significant effect on students' learning in a short period. Iranian students who have been studying in traditional teacher-centered classes for more than 12 years by sitting quite passively and listening to the teachers' lecturers find it not easy to adapt to the flipped classroom, as a student-centered method, in just a few sessions or even a semester. Probably, they need more time to embrace it more wholeheartedly. Third, another probable reason closely related to the second explanation is the students' insufficient engagement with the components of the new method. Despite their initial enthusiasm for this new teaching method, which was largely due to its novelty, it gradually became apparent that most of them had lost their zest and were not sufficiently and actively engaged in pre-class and in-class activities.

Regarding knowledge retention, the findings showed a statistically significant difference in favor of the flipped class. The results seem to be partly in accord with Noroozi et al. (2020) that examined the flipped classroom effect on Iranian EFL learners' acquisition of English conditionals. They reported that the

experimental group who attended a flipped class outperformed the control group in the delayed posttests administered two weeks after instruction. Even though Shatto et al.'s (2017) study was conducted in the field of nursing and its retention intervals were different, the results of this study are somewhat in line with the findings of their research, too. In contrast, the present study finding regarding knowledge retention differs from Bouwmeester et al.'s (2019) that revealed retention of knowledge tests after ten months showed similar outcomes for students in a flipped class and those in a traditional class. Morton and Colbert-Getz (2016) also demonstrated that flipped classroom students' ability to retain knowledge was equal to that of their counterparts who had attended a traditional class. However, their study showed that students in the flipped class appeared to be more successful in applying knowledge in a test administered at the end of the course.

The current study finding on knowledge retention could be attributed to several possible reasons. First, the demonstrated superiority of the flipped class over the traditional class to retain both explicit and implicit knowledge possibly lies in incorporating principles of active learning which has been suggested to enhance knowledge retention (Cherney, 2008; Semb & Ellis, 1994). Semb and Ellis (1994), for example, confirmed that instructional "strategies that more actively involved students in the learning process" resulted in higher differential retention (p. 277). Craik and Lockhart's (1972) "levels of processing theory" may also account for the higher retention of knowledge in the flipped classroom. They claim the deeper the information processing is, the longer the memory trace lasts. Craik (1973) defines depth as, "the meaningfulness extracted from the stimulus rather than in terms of the number of analyses performed upon it." (p. 48). Various meaningful activities performed under the heading of active learning in the flipped class can result in deep information processing and hence enhance knowledge retention. The second possible reason could be repetition which has been confirmed by numerous studies as a contributing factor (e.g., Fukuta, 2016; Kang, 2016). A flipped classroom can provide learners with multiple opportunities in the form of various activities and assignments to repeat content before, during, and after the class, and thus can improve learning and help them to retain knowledge longer. Testing effect or test-enhanced learning could be the third possible reason for the longer retention of knowledge in the experimental group. Studies carried out on the testing effect has

demonstrated that administering tests during the learning process augments learning and knowledge retention, the benefits of testing are superior to restudying the material, and this is more evident on delayed posttests (e.g., Butler, 2010; Schwieren et al., 2017). In-class and out-of-class quizzes in the flipped class may bring about a positive effect on retaining students' grammar knowledge.

Furthermore, as the results indicated, implicit grammar knowledge showed a higher retention rate in both groups. This probably means that the forgetting rates of the two types of knowledge are different and implicit knowledge seems to be more resistant to loss. This possibility has also been raised by many other studies (e.g., Reber, 1989; Tunney, 2003). Reber (1989), for example, suggests "implicit systems are robust in the face of disorders that are known to produce serious deficits in conscious, overt processes" (p. 232). Some other studies have argued that the robustness of implicit learning is less affected by divided attention (e.g., Prull et al., 2016) and interference (Tamayo & Frensch, 2007; Tunney, 2003). Tamayo and Frensch (2007) draw upon the multiple memory systems theory to explain why interference might result in different rates of memory loss and forgetting for explicit and implicit knowledge. They argue that different memory systems that might be affected differently by interference support explicit and implicit knowledge.

Conclusion and Implications

The results of this study suggest that offering explicit grammar teaching in the flipped classroom was as effective as explicit grammar instruction provided during the traditional class, that flipped class had a more favorable effect on the learners' grammar knowledge retention, and the forgetting rates of explicit and implicit knowledge are not the same and the latter is less susceptible to loss. By providing empirical evidence for the advantages of flipped classroom model and explicit instruction, the findings can help and ensure educational policymakers, language program designers, and language teachers to capitalize on explicit instruction for teaching grammar in various educational settings including flipped classes in which a portion of the instruction is moved out of class with no decline in the learners' achievement. This study may also offer some insights into the potentials of using technology in education in general and the role of technology-enhanced instruction in foreign language development in particular. The findings

can also make a contribution to blended learning by showing how advantages inherent in this model can affect learners' language development. In addition, the results show that delegating part of the responsibility of learning to students and employing active learning strategies can lead to genuine educational achievements. In particular, it can initiate a gradual movement towards student-centered education in the Iranian higher education context mainly dominated by teacher-centered instruction. This study can also contribute to the debate about the effectiveness of explicit instruction, as a type of FFI, and its contribution to the language learners' explicit and implicit L2 grammar knowledge.

Several limitations, however, were imposed on the findings of the study. First, the long-term effect of explicit teaching was tested after a four-week interval; longer intervals may lead to different results. Thus, future studies can extend the interval and examine longer retention of L2 forms and structures. Second, all the students were at the intermediate level of English proficiency. Therefore, further similar studies could include participants with higher or lower L2 proficiency levels. Third, this study focused on tertiary education students; primary or secondary school students can be selected in future studies as research samples. Fourth, probably several participants already had some knowledge of target grammar points taught in this study. This could possibly affect the results. Hence, it is recommended that an artificial grammar be used in future studies. Fifth, research has shown that explicit instruction might have different effects on different L2 forms. The reported results in this study are based on scores obtained from tests which included items on English passive voice, conditionals, and articles. Different results might have been obtained if each of these three forms or structures had been chosen as the only target form in the study. Finally, given the broad definition of the flipped class in the literature, it is not surprising that the concept has been implemented differently in various studies. Differences in the ways the flipped class concept is implemented is a potential source of variation in research results. Therefore, authors should describe all characteristics and particularities of the implemented flipped classes in sufficient detail to allow other researchers to compare the results or replicate their studies. Consequently, the findings of the present study cannot be generalized to other contexts in which the flipped class model has been implemented differently.

References

Afzali, Z., & Izadpanah, S. (2021). The effect of the flipped classroom model on Iranian English foreign language learners: Engagement and motivation in English language grammar. *Cogent Education, 8*(1). https://doi.org/10.1080/2331186X.2020.1870801

Akakura, M. (2012). Evaluating the effectiveness of explicit instruction on implicit and explicit L2 knowledge. *Language Teaching Research, 16*(1), 9-37. https://doi.org/10.1177/1362168811423339

Amiryousefi, M. (2019). The incorporation of flipped learning into conventional classes to enhance EFL learners' L2 speaking, L2 listening, and engagement. *Innovation in Language Learning and Teaching, 13*(2), 147-161. https://doi.org/10.1080/17501229.2017.1394307

Asaka, S., Shinozaki, F., & Yoshida, H. (2018). The effect of flipped classroom approach on EFL Japanese junior high school students' performances and attitudes. *International Journal of Heritage, Art, and Multimedia, 1*(3), 71-87. https://cutt.ly/qEaTxAG

Badrkhani, P. (2021). How a catastrophic situation turns into an exceptional opportunity: Covid-19 pandemic in Iran and challenges of online education for new English language educators. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2021.1956545

Bergmann, J., & Sams, A. (2012). *Flip your classroom: Reach every student in every class every day*. International Society for Technology in Education.

Bezzazi, R. (2019). Learning English grammar through flipped learning. *The Asian Journal of Applied Linguistics, 6*(2), 170-184. https://caes.hku.hk/ajal/index.php/ajal/article/view/597

Bouwmeester, R. A. M., de Kleijn, R. A. M., van den Berg, I. E. T., ten Cate, O. T. J., van Rijen, H. V. M., & Westerveld, H. E. (2019). Flipping the medical classroom: Effect on workload, interactivity, motivation and retention of knowledge. *Computers & Education, 139*(1), 118-128. http://dx.doi.org/10.1016/j.compedu.2019.05.002

Bowles, M. (2011). Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute? *Studies in Second Language Acquisition, 33*(2), 247-271. https://doi.org/10.1017/S0272263110000756

Bulut, C., & Kocoglu, Z. (2020). The flipped classroom's effect on EFL learners' grammar knowledge. *International Journal of Mobile and Blended Learning*, *12*(4), 69-84. https://bit.ly/2XxL8AX

Butler, A. C. (2010). Repeated testing produces improved transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1118-1133. https://doi.org/10.1037/a0019902

Cherney, I. D. (2008). The effects of active learning on students' memories for course content. *Active Learning in Higher Education, 9*(2), 152-171. https://doi.org/10.1177/1469787408090841

Cilliers, L., & Pylman, J. (2020). South African students' perceptions of the flipped classroom: A case study of higher education. *Innovations in Education and Teaching International*, *59*(3), 296-305. https://doi.org/10.1080/14703297.2020.1853588

Clark, K. R. (2015). The effects of the flipped model of instruction on student engagement and performance in the secondary mathematics classroom. *Journal of Educators Online, 12*(1), 91-115. http://files.eric.ed.gov/fulltext/EJ1051042.pdf

Craik, F. I. M. (1973). A levels of analysis view of memory. In P. Pliner, L. Kramer, & M. Alloway (Eds.), *Communication* and *affect: Language and thought* (pp. 45-65). Academic Press. https://doi.org/10.1016/B978-0-12-558250-6.50010-9

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X

De Graaff, R. (1997). The eXperanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition*, *19*(2), 249-276. https://doi.org/10.1017/S0272263197002064

DeKeyser, R. (2003). Explicit and implicit learning. In C. Doughty, & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313-348). Blackwell. http:/dx.doi.org/10.1002/9780470756492.ch11

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*. Dover.

Ellis, N. C. (2017). Implicit and explicit knowledge about language. In J. Cenoz, D. Gorter, & S. May (Eds.), *Language awareness and multilingualism. Encyclopedia of language and education* (3rd ed., pp. 113-124). Springer.

Ellis, R. (2002). Does form-focused instruction affect the acquisition of implicit knowledge? A review of the research. *Studies in Second Language Acquisition, 24*(2), 223-236. https://doi.org/10.1017/S0272263102002073

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning, 54*(2), 227-275. https://doi.org/10.1111/j.1467-9922.2004.00255.x

Ellis, R. (2008). Investigating grammatical difficulty in second language learning: Implications for second language acquisition research and language testing. *International Journal of Applied Linguistics, 18*(1), 4-22. https://doi.org/10.1111/j.1473-4192.2008.00184.x

Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit*

knowledge in second language learning, testing and teaching (pp. 3-25). Multilingual Matters. https://doi.org/10.21832/9781847691767-003

Ellis, R., & Loewen, S. (2007). Confirming the operational definitions of explicit and implicit knowledge in Ellis (2005): Responding to Isemonger. *Studies in Second Language Acquisition, 29*(1), 119-126. https://doi.org/10.1017/S0272263107070052

Fathi, J., & Rahimi, M. (2020). Examining the impact of flipped classroom on writing complexity, accuracy, and fluency: A case of EFL students. *Computer Assisted Language Learning*, *35*(7), 1668-1706. https://doi.org/10.1080/09588221.2020.1825097

Francis, J. (2017). *The effects of technology on student motivation and engagement in classroom-based learning* [Doctoral dissertation, University of New England]. https://dune.une.edu/theses/121.

Fukuta, J. (2016). Effects of task repetition on learners' attention orientation in L2 oral production. *Language Teaching Research, 20*(3), 321-340. https://doi.org/10.1177/1362168815570142

Graham, K. L., Cohen, A., Reynolds, E. E., & Huang, G. C. (2019). Effect of a flipped classroom on knowledge acquisition and retention in an internal medicine residency program. *Journal of graduate medical education*, *11*(1), 92-97. https://doi.org/10.4300/JGME-D-18-00536.1

Haynie, W. J. (2007). Effects of test taking on retention learning in technology education: A meta-analysis. *Journal of Technology Education*, *18*(2), 24-36. https://doi.org/10.21061/jte.v18i2.a.2

Housen, A., & Pierrard, M. (2005). Investigating instructed second language acquisition. In A. Housen, & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp.1-27). Mouton de Gruyter. https://doi.org/10.1515/9783110197372

Hulstijn, J. H., & de Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review*, *11*, 97-112. https://hdl.handle.net/11245/1.428191

Izadpanah, S., & Afzali, Z. (2021). The effect of the flipped classroom model on critical thinking and performance of Iranian EFL learners in learning grammar. *Journal of Language Horizons*, 6(1), 183-204. https://doi.org/10.22051/lghor.2021.32997.1362

Kang, N. (2015). The comparison between regular and flipped classrooms for EFL Korean adult learners. *Multimedia- Assisted Language Learning, 18*(3), 41-72. https://doi.org/10.15702/MALL.2015.18.3.41

Kang, S. H. K. (2016). Spaced repetition promotes efficient and effective learning: Policy implications for instruction. *Policy Insights from the Behavioral and Brain Sciences*,

3(1), 12-19. https://doi.org/10.1177/2372732215624708

Kang, E. Y., Sok, S., & Han, Z. H. (2019). Thirty-five years of ISLA on form-focused instruction: A meta-analysis. *Language Teaching Research, 23*(4), 428-453.

Khezrlou, S. (2021). Effects of timing and availability of isolated FFI on learners' written accuracy and fluency through task repetition. *The Language Learning Journal, 49*(5), 568-580. https://doi.org/10.1080/09571736.2019.1656765

King, A. (1993). From sage on the stage to guide on the side. *College Teaching, 41*(1), 30-35. https://doi.org/10.20429/ijsotl.2014.080104

Lee, A. H. (2021). The effects of proactive form-focused instruction and individual differences on second language acquisition. *Language Teaching Research, 25*. https://doi.org/10.1177/13621688211051032

Lee, G., & Wallace, A. (2018). Flipped learning in the English as a foreign language classroom: Outcomes and perceptions. *TESOL Quarterly, 52*(1), 62-84. https://doi.org/10.1002/tesq.372

Li, S., Ellis, R., & Zhu, Y. (2019). The associations between cognitive ability and L2 development under five different instructional conditions. *Applied Psycholinguistics*, *40*(3), 693-722. https://doi.org/10.1017/S0142716418000796

Lichtman, K. (2016). Age and learning environment: Are children implicit second language learners? *Journal of Child Language, 43*(3), 707-730. https://doi.org/10.1017/S0305000915000598

Lindseth, M. (2016). The effects of form-focused instruction on the acquisition of subject-verb inversion in German. *Foreign Language Annals, 49*(1), 10-22. https://doi.org/10.1111/flan.12174

Liu, C., Sands-Meyer, S., & Audran, J. (2019). The effectiveness of the student response system (SRS) in English grammar learning in flipped English as a foreign language (EFL) class. *Interactive Learning Environments*, *27*(8), 1178-1191. https://doi.org/10.1080/10494820.2018.1528283

Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach*. Benjamins. https://doi.org/10.1075/lllt.18

McLaughlin , E. J. (2018). Flipped classrooms, by design. *Medical Education, 52*(9), 887-888. https://doi.org/10.1111/medu.13654

Morton, D. A., & Colbert-Getz, J. M. (2016). Measuring the impact of the flipped anatomy classroom: The importance of categorizing an assessment by Bloom's taxonomy. *Anatomical Sciences Education, 10*(2), 170-175. https://doi.org/10.1002/ase.1635

Noroozi, A., Rezvani, E., & Ameri-Golestan, A. (2020). The effect of flipped classrooms on L2 learners' development and retention of grammatical knowledge. *Turkish Online*

Journal of Distance Education, *21*(4), 14-30. https://doi.org/10.17718/tojde.803344

Norris, J. M., & Ortega, L. (2000). Effectiveness in L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*(3), 417-528. https://doi.org/10.1111/0023-8333.00136

Nourinezhad, S., Hadipourfard, E., & Bavali, M. (2022). The effect of flipped learning on English writing performance and self-efficacy of Iranian medical students. *Journal of Language Horizons*, *6*(1), 161-182. https://doi.org/10.22051/lghor.2021.34132.1409

O'Flaherty, J., & Phillips, C. (2015). The use of flipped classrooms in higher education: A scoping review. *Internet and Higher Education, 25*, 85-95. https://doi.org/10.1016/j.iheduc.2015.02.002

Prull, M. W., Lawless, C., Marshall, H. M., & Sherman, A. T. K. (2016). Effects of divided attention at retrieval on conceptual implicit memory. *Frontiers in Psychology, 7.* https://doi.org/10.3389/fpsyg.2016.00005

Ranta, L., & Lyster, R. (2018). Form-focused instruction. In P. Garrett & J. Cots (Eds.), *The Routledge handbook of language awareness* (pp. 40-56). Routledge.

Rassaei, E. (2017). Video chat vs. face-to-face recasts, learners' interpretations and L2 development: A case of Persian EFL learners, *Computer Assisted Language Learning, 30*(1-2), 133-148. https://doi.org/10.1080/09588221.2016.1275702

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219-235. https://doi.org/10.1037/0096-3445.118.3.219

Roehling, P. V. (2018). *Flipping the college classroom: An evidence-based guide*. Palgrave. https://doi.org/10.1007/978-3-319-69392-7

Saito, K. (2013). Reexamining effects of form-focused instruction on L2 pronunciation development. *Studies in Second Language Acquisition*, *35*(1), 1-29. https://doi.org/10.1017/S0272263112000666

Sanz, C., & Morgan-Short, K. (2004). Positive evidence versus explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning*, *54*(1), 35-78. https://doi.org/10.1111/j.1467-9922.2004.00248.x

Schwieren, J., Barenberg, J., & Dutke, S. (2017). The testing effect in the psychology classroom: A meta-analytic perspective. *Psychology Learning & Teaching, 16*(2), 179-196. https://doi.org/10.1177/1475725717695149

Semb, G. B., & Ellis, J. A. (1994). Knowledge taught in school: What is remembered? *Review of Educational Research, 64*(2), 253-286. https://doi.org/10.3102/00346543064002253

Shatto, B., L'Ecuyer, K., & Quinn, J. (2017). Retention of content utilizing a flipped classroom approach. *Nursing Education Perspectives, 38*(4), 206-208.

https://doi.org/10.1097/01.NEP.0000000000000138

Spada, N. (2011). Beyond form-focused instruction: Reflections on past, present and future research. *Language Teaching, 44*(2), 225-236. https://doi.org/10.1017/S0261444810000224

Tamayo, R., & Frensch, P. A. (2007). Interference produces different forgetting rates for implicit and explicit knowledge. *Experimental Psychology, 54*(4), 304-310. https://doi.org/10.1027/1618-3169.54.4.304

Toni, A., & Hassaskhah, J. (2018). Form-focused skill training in EFL speaking pedagogy: Empirical counterevidence to trade-off hypothesis. *Revista de Educación*, *380*, 9-41. https://doi.org/10.4438/1988-592X-RE-2017-380-371

Trahey, M., & Spada, N. (2020). Focusing on language and content with adolescent English language learners in the mainstream classroom. *Canadian Modern Language Review*, *76*(3), 218-242. https://doi.org/10.3138/cmlr-2019-0009

Tunney, R. J. (2003). Implicit and explicit knowledge decay at different rates: A dissociation between priming and recognition in artificial grammar learning. *Experimental Psychology, 50*(2), 124-130. https://doi.org/10.1026/1618-3169.50.2.124

Turan, Z., & Akdag-Cimen, B. (2020). Flipped classroom in English language teaching: A systematic review. *Computer Assisted Language Learning, 33*(5-6), 590-606. https://doi.org/10.1080/09588221.2019.1584117

Vaezi, R., Afghari, A., & Lotfi, A. (2019). Flipped teaching: Iranian students' and teachers' perceptions. *Applied Research on English Language, 8*(1), 139-164. https://doi.org/10.22108/are.2019.114131.1382

Vitta, J. P., & Al-Hoorie, A. H. (2020). The flipped classroom in second language learning: A meta-analysis. *Language Teaching Research,* 1-25. https://doi.org/10.1177/1362168820981403

Webb, M., & Doman, E. (2016). Does the flipped classroom lead to increased gains on learning outcomes in ESL/EFL contexts? *CATESOL Journal, 28*(1), 39-67. http://www.catesoljournal.org/wp content/uploads/2016/06/CJ28.1_webb.pdf

Woolfolk, A., & Margetts, K. (2016). *Educational psychology* (4th Ed.). Pearson Education Australia.

Xu, J. & Li, C. (2021). The effects of the timing of form-focused instruction on EFL learners' learning of difficult and easy grammatical features: A comparative study. *System, 101*. https://doi.org/10.1016/j.system.2021.102612

Xu, H., & Lyster, R. (2014). Differential effects of explicit form-focused instruction on morphosyntactic development. *Language Awareness*, *23*(1-2), 107-122. https://doi.org/10.1080/09658416.2013.863899

Zainuddin, Z., & Perera, C. J. (2019). Exploring students' competence, autonomy and relatedness in the flipped classroom pedagogical model. *Journal of Further and Higher Education*, *43*(1), 115-126. https://doi.org/10.1080/0309877X.2017.1356916

Journal of Language Horizons

Scientefic Quarterly Journal of Language Horizons, *Alzahra University*
Volume 7, Issue 1, Spring 2023 (Serial No. 15)

The Impact of Self-Regulated Strategy Development on Intermediate EFL Learners' Capability in Self-Regulating Reading

Research Article
pp. 135-157

Mahboobeh Morshedian[1]

Received: 2021/07/31 Accepted: 2022/04/23

Abstract

Academic self-regulation is a process in which learners use cognitive, affective, behavioral, and motivational feedback to adjust or modify their behaviors and strategies to reach their objectives. Similarly, self-regulated learning (SLR) is the process of defining objectives, planning strategically, choosing and utilizing strategies, monitoring one's efficiency, and self-evaluating oneself. Accordingly, this research explored the influence of self-regulation teaching based on Think Before, While, and After reading (TWA) strategy with self-regulated strategy development (SRSD) on intermediate English as a foreign language (EFL) learners' self-regulation of their EFL reading of expository texts. SRSD for TWA strategy was implemented in the EFL reading sessions of the intact experimental group, but the intact control group was given routine EFL reading instruction, i.e., to read the texts and answer its comprehension questions. The self-regulation data was collected utilizing the adjusted Motivated Strategies for Learning Questionnaire Parametric (MSLQ) before and after the instruction. Through one-way analysis of the covariance, it was shown that self-regulation instruction based on TWA with SRSD could foster self-regulatory reading skills of EFL learners. These findings can urge teachers to teach EFL readers SRSD for TWA to improve their ability to self-regulate their EFL reading process.

Keywords: self-regulation, strategy development, reading, self-regulated learning (SLR), intermediate EFL learners

[1] Assistant Professor, Department of English Translation, Faculty of Humanities, Tolou-e Mehr Non-profit Institute of Higher Education, Qom, Iran. m.morshedian@tolouemehr.ac.ir

DOI: 10.22051/lghor.2022.37106.1533
DOR: 20.1001.1.2588350.2023.7.1.6.4

Introduction

As Second Language (L2) reading leads to not only learning but also enjoyment (Nation, 2009), it is of paramount importance in the second language (L2) acquisition. Mori (2004) maintains that reading is also an essential and most probably a crucial skill for learners in the English as a Foreign Language (EFL) context. This necessitates the development of necessary reading strategies in EFL learners (Grabe, 2009, 2014), including self-regulation or Self-Regulated Learning (SRL) strategies which can enhance reading in both First Language (L1) (e.g., Housand & Reis, 2008; Swalander & Taube, 2007) and L2/EFL (e.g., Mbato, 2013; Morshedian et al. 2017).

Academic self-regulation is a process in which learners use cognitive, affective, behavioral, and motivational feedback to adjust or modify their behaviors and strategies in order to achieve their goals. Similarly, SLR is the degree to which students can define goals, make a strategic plan, choose and utilize strategies, and self-monitor and self-evaluate their effectiveness and themselves (Zimmerman, 2008). On the one hand, teachers can explicitly teach and model self-regulation and its strategies (Zimmerman, 2002). On the other hand, skillful readers are regarded as self-regulating and active learners (Harris & Pressley, 1991, as cited in Woolley, 2011). Hence, a vital dimension of being a skillful reader is self-regulating one's reading process (Butler, 2002, as cited in Woolley, 2011).

Among the self-regulation models that were applied to L1 reading (e.g., Mason et al., 2006) and can be used in EFL reading (e.g., Hamoulah Mardani & Afghari, 2017; Roohani & Asiabani, 2015) is Self-Regulated Strategy Development (SRSD; Harris & Graham, 1996). It has been originally developed to assist students who encounter significant writing difficulties, but because it is a teaching model that combines the teaching of self-regulation processes and strategy instruction (Mason, 2004), it can be applied to other language skills such as reading. It is aimed at helping students become independent, fluent, and goal-directed learners (Mason, 2004). On the whole, in the L1 context, Graham, Harris, their colleagues, and other researchers have validated SRSD with students struggling with reading in more than 25 research studies conducted in 20 years (Graham & Harris, 2003, as cited in Mason, 2004).

SRSD involves six stages, including modeling, guided (collaborative)

practice, and pair/independent practice. Mason (2004) created and implemented it for TWA (i.e., think before, while, and after reading), which is a technique fostering reading (Mason et al., 2006). However, the studies which have been conducted on L2 reading self-regulation (e.g., Mbato, 2013; Morshedian et al., 2017) have not applied SRSD (Harris & Graham, 1996) for Mason's (2004) TWA strategy on EFL reading, and those few that did so (e.g., Hamoulah Mardani & Afghari, 2017; Roohani & Asiabani, 2015; Roohani et al., 2016; Roohani et al., 2017) either focused on the argumentative texts or excluded TWA from their studies, or did not explore the resulting self-regulation ability of EFL learners while reading. Thus, the following research question was formed:

Does self-regulation teaching based on TWA with SRSD significantly influence the intermediate EFL learners' self-regulation of reading?

Review of Literature

SRDS and TWA

SRSD (Harris & Graham, 1996) is an instruction model that brings explicit teaching of self-regulation together with strategy teaching. SRSD was initially created by Harris and Graham (1996) to improve L1 writing in students with learning disabilities or low-achieving students. It has these stages; (a) Create background: prior knowledge is activated and discussed to make sure students have knowledge and skill necessary for reading tasks; (b) Talk about It: the teacher discusses how the strategy use assists reading, (c) Model It: the teacher models cognitively (i.e., his thinking out loud while putting the strategy into practice), (d) Memorize It: the teacher and students commit the strategy to their memory, (e) Back It up: students and the teacher practice it cooperatively and students practice it with partners, and (f) Independent Performance: students practice it independently **(**Harris & Graham, 1999; Mason et al., 2006).

On the other hand, TWA strategy (think before, while, and after reading), which can foster reading, can be implemented through SRSD (Mason et al., 2006). It is a procedure that includes cognitive strategies in a framework for active engagement in text and can be taught through the SRSD instruction (Mason et al., 2006). In TWA, students learn strategies that can be utilized before, while, and after reading a text. Before they start reading, they learned to find the writer's purposes,

think about their knowledge, and decide about their future learning. Implemented through SRSD, TWA teaching starts with modeling, then comes guided (collaborative) practice, and the final phase is pair or independent practice (Mason, 2004; Mason et al., 2006) (see Table 1).

Table 1

TWA Strategy

(T) think before reading (on)	(W) think while reading (on)	(A) think after reading (on)
The writer's intention	Reading speed	The main idea
What you know	Linking what you know	Summarizing information
What you want to learn	Rereading parts	What you learned

Empirical Studies on SRDS for TWA

Empirical studies investigating TWA within SRSD showed that it could improve L1 reading in struggling readers (e.g., Hedin et al., 2011; Mason et al., 2006; Mason et al., 2012; Mason et al., 2013). Mason (2004), for instance, compared explicit SRSD for TWA with reciprocal questioning strategies in thirty-two fifth-grade students who had difficulty in L1 reading. Both groups showed improvement in their reading ability; however, the TWA intervention was more effective than the reciprocal questioning groups. Mason et al. (2013) also provided SRSD teaching for the TWA for low-achieving students and found out about students' knowledge about the self-regulation techniques taught and learned. Reviewing some studies on SRSD for TWA in both L1 writing and reading, Mason (2013) also reported that students' reading of L1 expository texts improved by SRSD for TWA.

In the L2 and EFL contexts, Hedin et al. (2011) examined the impact of teaching TWA on L2 learners having attention-related and comprehension problems. Through scaffolded support and self-monitoring, their reading improved, and they could regulate their strategy use. Roohani and Asiabani (2015) found out that SRSD through TWA instruction could improve the EFL learners' reading of argumentative texts and significantly enhanced their metacognitive awareness. Interestingly, Hamoulah Mardani and Afghari's (2017) research showed that SRSD could enhance EFL learners' reading of argumentative texts, but it did not significantly influence their metacognitive knowledge. However, these studies neither mixed SRSD for TWA nor explored the resulting self-regulation ability of EFL learners while

reading. In addition, Mason et al. (2006) maintain that TWA with SRSD is an excellent option for teachers to begin, develop, and enhance learners' competence in L1 expository reading. Nevertheless, as mentioned above, no study has yet applied SRSD (Harris & Graham, 1996) for Mason's (2004) TWA strategy on EFL reading, using expository texts and exploring the resulting self-regulation ability of EFL learners while reading, so this research delved into how such specific self-regulation teaching could enhance self-regulation ability in the EFL expository reading.

Method

Participants

The participants in this research study were 40 female Iranian EFL learners placed in two intact reading classes and taught by the researcher in an English language institute. The age range of the participants differed between 19 and 31 (M = 23.30, SD = 6.14). Under the institute policy, they were placed in two intermediate classes of 20 students based on results obtained from a Preliminary English Test (PET) (Hashemi &Thomas, 2009), and *Inside Reading 2* (2nd ed., Zwier, 2016) constituted their reading material. Each class was randomly chosen as the control and experimental groups. That is, the experimental class learned self-regulation as aimed at EFL reading based on SRSD for TWA, and the control class was taught based on routine EFL reading teaching method, i.e., to read the reading selections and do their comprehension exercises.

The Instructional Materials

The reading texts of intermediate classes were chosen from *Inside Reading 2* (2nd ed., Zwier, 2016). The readability indices of two texts from the textbook, calculated through Flesch readability ease (FRE) (Taylor & Weir, 2012), displayed scores ranging from 58 to 64. This indicates standard/average texts to read (DuBay, 2006) that were at the appropriate level for the intermediate participants. The reasons why the teacher worked on the expository reading texts are as follows. First, the expository genre includes various text types and entails using many reading strategies simultaneously (Gersten et al., 2001). Second, empirically text genre does not have an important part in the L2/EFL learners' reading capability (Allen et al., 1988). It is also noteworthy that in both groups the students merely did the

comprehension exercises of the chosen reading texts and did not have to do reading exercises either preceding or following the reading selections.

Instruments

As students' self-regulation ability could be uncovered from self-reported behavior in questionnaires (e.g., Entwistle, 1988, as cited in Boekaerts & Corno, 2005), Pintrich et al.'s (1993) Motivated Strategies for Learning Questionnaire (MSLQ), which includes motivation and learning strategies scales was adapted to reading and was utilized to gauge reading self-regulation. In adapting it, its wording was altered to a small extent so that every sentence seemed relevant to reading. As an example, the sentence, "If I try hard enough, then I will understand the *course* material," was changed to this senetnce, "If I try hard enough, then I will understand the *reading* material."

The internal consistency coefficients reported for MSLQ subscales differed from .52 to .92 (Winne & Perry, 2000). Pintrich et al.'s (1993) also proved the content, predictive, and construct validity of the scale. Through confirmatory factor analyses, the developers tested the use of the theoretical model of the MSLQ and specified which items were allocated to which factor. Likewise, from the close relationship between a domain of theory and the scale's items, they established its content validity. The scales in the Learning Strategy section and those in the Motivation section also showed an empirically validated framework for evaluating students' learning strategies and motivation. They also investigated its predictive validity in terms of higher education academic achievement. Most scales in the Motivation and the Learning Strategy section showed significant correlations with final course grades. In this research, the reliability estimates of the pretest and the posttest were found to be $\alpha = .72$ and $\alpha = .77$, respectively. However, the adapted questionnaire was not validated due to practical restrictions, which is considered a limitation of this study.

The Pintrich's (2000) self-regulation model which has some characteristics in common with TWA forms the basis of this questionnaire (Mason et al., 2006). In other words, it has almost three phases roughly corresponding to those TWA (Mason et al., 2006). His model includes the following stages: 1) activation, planning, and forethought 2) monitoring, 3) control, and 4) reflection and reaction;

and TWA (Mason et al., 2006) involves "think before, while, and after reading." Moreover, MSLQ considers self-regulation an aptitude (Winne & Perry, 2000) which is enduring, at least during research that may take a few weeks (Boekaerts et al., 2000). Hence, a single measurement puts different pieces of information on SRL together and forms a total, predicting a student's self-regulation ability (Winne & Perry, 2000) (see Appendix A).

Procedure and Data Collection

Before the treatment, participants were given the Pintrich et al.'s (1993) MSLQ, as adjusted to reading to gauge their capatility to practice reading self-regulation. Then SRSD (Harris & Graham, 1996) was implemented in the experimental group for TWA strategy for comprehension of expository reading texts (Mason, 2004). The author (i.e., the teacher) had three briefing meetings with learners before the treatment administration and put SRSD for TWA into practice on sample practice reading tasks so that they became thoroughly familiar with the treatment. Following Mason et al. (2006), in the before-reading stage of TWA, she taught students to make background knowledge active by guessing the writer's purpose, their knowledge of the topic, and their future learning about it through reading this text (Ogle, 1989, as cited in Mason et al., 2006). During reading, students were urged to ponder on their reading pace, connect what they know, and read parts of text again (Graves & Levin, 1989, as cited in Mason et al., 2006). After reading, students first learned to find main ideas utilizing the RAP strategy (i.e., reading the paragraph, asking themselves what sentence in the paragraph includes its main idea, paraphrasing the main idea) (Ellis & Graves, 1990, as cited in Mason et al., 2006). Then, they summarized or paraphrased the reading passage, utilizing Brown and Day's (1983, as cited in Mason et al., 2006) summarization strategy (i.e., deleting trivial and redundant information, replacing a list of words or actions with super-ordinate words, selecting or making a topic sentence). Finally, students were asked to retell the passage orally, with the teacher's support when necessary.

All the above-mentioned procedure of TWA (Mason, 2004) was implemented through the framework of SRSD (Harris & Graham, 1996) based on Hedin et al. (2011) and Mason et al. (2006). In other words, the researcher began instruction with modeling, proceeded with guided (collaborative) practice, and

finished it with pair or independent practice as follows:

Discussion of Strategy. The teacher discussed TWA use and how it can help reading, fostered pre-skills necessary for reading expository texts, and explained each step of TWA.

Modeling. The students observed and interacted with the teacher as strategy use was modeled and demonstrated by her. Moreover, she used self-sentences during reading to support and guide the TWA process. As an example, she may say before reading, "It is easy to understand this text since I am sure the stages in TWA will help me comprehend the text. What should I do first?" Students followed the teacher and wrote self-statements they could utilize before, while, and after reading.

Guided (Collaborative) Practice Leading to Memorization of Strategy. In collaborative practice, the teacher and students worked together to put into practice the phases of TWA. The teacher encouraged students to utilize their self-statements when reading and to monitor strategy steps that were completed. The teacher observed student progress. Later, when students mastered the use of strategies, the teacher helped them merely when there was a problem in the practice of strategy. Helping students memorize the strategy steps and complete them, the teacher urged them to reflect on the reponses after reading, reminding them of putting phases into practice.

Pair (Independent) Practice. Pair practice was put into practice after students became skillful in using TWA with teacher support. At first, student pairs practiced doing each phase, monitored their performance, and reported the outcomes of their reading to the instructor, with the instructor helping them when necessary. This practice was repeated until students could go through the stages of the strategy by themselves without the teacher or classmate's help.

In the control group, the students studied the same expository reading passages as those read in the experimental group but based on the routine method. In other words, they answered pre-reading questions, read the text, summarized or paraphrased it, and answered post-reading questions. Then after the treatment, participants were given Pintrich et al.'s (1993) MSLQ, as adjsuted to reading, to gauge their EFL reading self-regulation.

Results

Preliminary checks for the one-way ANCOVA test involve the examination of normality, the reliability of the covariate (the pretest in this study), the linearity of the relationship between the covariate and the dependent variable (i.e., the posttest), and the homogeneity of regression slopes (Pallant, 2005). Table 2 below depicts the results of the normality test for MSLQ.

Table 2

Test of Normality for MSLQ

		Kolmogorov-Smirnova			Shapiro-Wilk		
	group	Statistic	df	*p*	Statistic	df	*p*
Post-MSLQ	experimental	.075	20	.200*	.959	20	.524
	control	.136	20	.200*	.945	20	.299

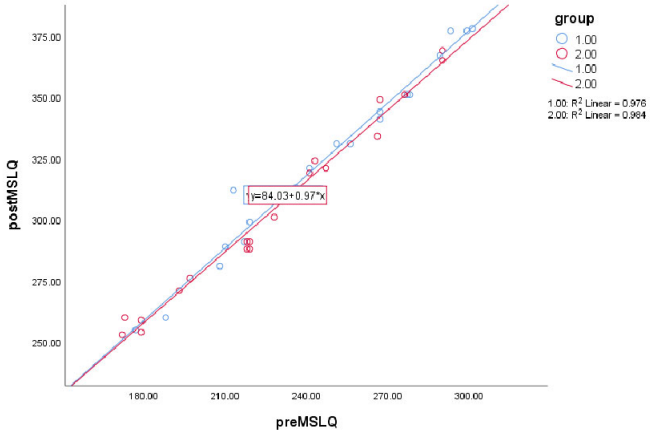
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

As Table 2 suggests, no violation of the normality assumption was detected (p >.05). Next, the reliability of the pretest was found to be α = .71. After that, the linearity assumption was examined. The reliability (i.e., internal consistency) of the pretest as calculated was found to be α =.72. Next, the linearity of the relationship between the covariate and the dependent variable was tested as can be observed in Figure 1.

Figure 1

Test of Linearity for MSLQ



The linear relationship between the posttest and the pretest (experimental groups R^2 Linear = .976 & control groups R^2 Linear = .984) can be seen in Figure 1. Next, the homogeneity of regression was tested, the result of which can be observed in Table 3.

Table 3

Test of Homogeneity of Regression Slopes for MSLQ

Source	Type III Sum of Squares	df	Mean Square	F	*p*
Corrected Model	55614.710a	3	18538.237	641.915	.000
Intercept	6400.130	1	6400.130	221.614	.000
group	2.099	1	2.099	.073	.789
Pre-MSLQ	51554.328	1	51554.328	1785.147	.000
group * pre-MSLQ	6.428	1	6.428	.223	.083
Error	1039.665	36	28.880		
Total	3997355.000	40			
Corrected Total	56654.375	39			

a. R Squared = .982 (Adjusted R Squared = .980)

Table 3 shows that no significant interaction was detected between groups and the pretest, indicating homogeneity of the regression slopes F (1, 36) = 6.42, p > .05. Table 4 below depicts the results of variance homogeneity test for MSLQ.

Table 4

Test of Variance Homogeneity for MSLQ

F	df1	df2	*p*
.020	1	38	.889

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

a. Design: Intercept + pre-MSLQ + group

Table 4 suggests that the error variance was equal across groups, and there was no violation of the assumption, F (1, 37) = .020, p > .05. The results of the preliminary checks for the MSLQ revealed that the one-way ANCOVA test could be used to compare the differences of groups in the posttest when pretest scores were

taken as a covariate (Pallant, 2005). The standard deviations and means of the control and experimental classes on MSLQ questionnaire are shown in Table 5 below.

Table 5

Descriptive Statistics of MSLQ

group	Mean	Std. Deviation	N
experimental	323.9	37.5	20
control	303.8	36.8	20
Total	313.8	38.1	40

Table 5 above shows a higher mean score in the posttest in the experimental class than the control class. Table 6 below depicts the one-way ANCOVA results for EFL reading self-regulation via MSLQ by Groups.

Table 6

One-way ANCOVA: Test of Self-regulation of EFL Reading through MSLQ by Groups

Source	Type III Sum of Squares	Df	Mean Square	F	*p*	η2
Corrected Model	55608.281a	2	27804.141	983.424	.000	.982
Intercept	6434.104	1	6434.104	227.572	.000	.860
Pre-MSLQ	51548.056	1	51548.056	1823.238	.000	.98
group	44.587	1	44.587	0.577	.017	.641
Error	1046.094	37	28.273			
Total	3997355.000	40				
Corrected Total	56654.375	39				

In Table 6, the one-way ANCOVA results on the MSLQ data depicts significant discrepancy between the learners' performance in the control group (M = 303.8, SD = 36.8) and the experimental one (M = 323.9, SD = 37.5) on the posttest, F (1,37) = 0.57, p = .017, η2 = .64.

Discussion

The results revealed that self-regulation teaching based on TWA (Mason, 2004) with SRSD (Harris & Graham, 1996) could foster the self-regulatory reading

skills of EFL learners. Considering the importance of EFL and L2 reading (Mori, 2004; Nation, 2009) and the fact that reading proficiency involves self-regulation (Woolley, 2011), the findings of this research both give us empirical evidence for TWA with SRSD and prove its application to EFL reading, which has not been dealt with in research on self-regulation of EFL and L2 reading through SRSD (e.g., Hamoulah Mardani &Afghari, 2017; Hedin et al., 2011; Roohani and Asiabani, 2015). Of course, the findings of Roohani and Asiabani (2015) and Roohani et al. (2016) that SRSD with TWA instruction could improve the EFL participants' metacognition which is a major component of self-regulation (Zimmerman, 2002, 2008) could be extended in the present study.

The findings of this research could not only empirically proved the successful application of TWA with SRSD on EFL expository texts but also prove that the EFL reading can be effectively regulated through it. Moreover, this research could experimentally extend Mason et al.'s (2006) to the EFL context; he maintains that SRSD for TWA helps readers to actively and consciously participate in the reading process. The results of this research which was the first to explore SRSD TWA on EFL expository texts also chime with Paris and Paris' (2001) support for teaching how to use SRL processes and Woolley's (2011) assertion that independent readers can be those who are taught self-regulation.

Similarly, the results have been consistent with those gained in the studies (e.g., Housand & Reis, 2008; Perry, 1998) delineating the teachers' main part in effectively helping learners in order to enhance their self-regulation ability in reading. This is congruent with theoretical assertions on teachers's role in fostering learners' self-regulation ability not only in academic subjects (Moos & Ringdal, 2012) but also in L1 reading (Paris & Paris, 2001). In addition, it highlights how Souvignier and Mokhlesgerami (2006) urged teachers to help learners in making use of reading strategies in a self-regulatory manner and how Butler (2002, as cited in Woolley, 2011) maintained that the crucial dimension of becoming a skilled reader is the capability to self-regulate one's reading process.

Likewise, the findings not only confirm that self-regulation of reading through TWA within SRSD which is teachable in the L1 context (Hedin et al., 2011; Mason, 2004; Mason et al., 2006; Mason et al., 2012; Mason, 2013; Mason et al., 2013) can be taught in the EFL context. In addition, they substantiate what Chamot

(2014) asserted about the part self-regulation plays in overcoming L2 reading difficulties and what Graham and Harris (2007) stated about how SRSD instruction could offer tangible models for what has to occur in the mind. The obtained results also corroborated those that showed how L2/EFL readers became self-regulated via other teaching plans (e.g., Mbato, 2013) and how other SRL models could be successfully taught L2 and EFL readers (e.g., Morshedian, et al., 2017).

On the one hand, Pintrich's (2000) self-regulation model, i.e., the basis of MSLQ (Pintrich et al., 1993) shares some features with TWA, i.e., it has almost three phases roughly corresponding to those TWA (Mason, 2004). On the other hand, there are three main objectives for SRSD teaching; helping learners develop higher-order cognitive processes entailed in academic setting, assisting them to foster self-regulatory strategies to manage and monitor their academic performance, and fostering positive attitudes in them about themselves as students and learning. Hence, these results seem worthwhile and interesting because the obtained results can be traced back in the experimental class's going through SRSD teaching in several phases; they practiced the TWA during these phases to uncover the text meaning via reflection. In the thinking-before-reading phase, they learned thinking about their knowledge and deciding what they are going to acquire. In the thinking-while-reading phase, they monitored their reading pace and connected their background knowledge to the reading content. In the thinking-after-reading phase, they summarized details, found the main idea, and evaluated what they have read.

Conclusion and Pedagogical Implications

The results of this research support the conclusion that SRSD for TWA can be effectively carried out in EFL reading classes to help learners self-monitor and self-evaluate while reading, which is regarded equal to reading competence (Butler, 2002, as cited in Woolley, 2011). In other words, it can be concluded that implementing TWA through the SRSD is effective in supporting EFL learners to become independent EFL readers and maintain their EFL reading skills. Likewise, in line with Graham & Harris (2007) assertion that in SRSD teaching the teacher will have students collaborate in finding the useful strategy and will pave the ground for students to utilize it, they indicate that not only EFL reading teachers but also, in the first place, syllabus designers and material developers should pay due attention

to SRSD self-regulatory procedure. That is, EFL redaing teachers are urged to utilize TWA through SRSD in their classes to assist students to practice reading self-regulation. From another perspective, neglecting the strategies' vital role in EFL reading particularly that of TWA through SRSD, makes students passive readers who cannot set attainable and realistic goals, self-regulate and self-evaluate and become proficient, independent readers. This study also contributes the field through its detailed procedure section, SRSD for TWA was introduced to EFL educators as a practical model to foster the self-regulation of their students.

Nonetheless, the limitations of this study should be acknowledged. This study investigated SRSD for TWA, and MSLQ (Pintrich et al., 1993), which roughly matches TWA, was utilized in this study because no measure has yet been developed to evaluate the outcome of implementing SRSD for TWA. Nevertheless, it would be worthy to carry out further research to further examine SRSD for TWA utilizing other SRL instruments or qualitatively by such instruments as think-aloud protocols or interviews. Also, the participants in the study were only female due to practical reasons. Thus, future studies are to be conducted to probe into the probable gender-related effects on the outcome of practicing self-regulation of EFL reading to generalize the results to males because the literature is inconclusive concerning the role of gender in self-regulation ability (Pintrich & Zusho, 2007).

Still, another limitation of this study seems to be no validation of MSLQ as adapted to the reading and the use of intact classes in the study because of the practical restrictions. Hence, future research can be conducted not only after validating MSLQ as adapted to the reading but also with the aim of delving into the impact of SRSD for TWA training on the reading ability of EFL learners, their motivation and attitude to read in English. The only delimitation of the study seems to be the inclusion of EFL expository texts in the research because they include different text types and involve using many reading strategies at the same time (Gersten et al., 2001).

References

Allen, E. D., Bernhardt, E. B., Berry, M. Th., & Demel, M. (1988). Comprehension and text genre: An analysis of secondary school foreign language readers. *The Modern Language Journal, 72*(2), 163-172. https://doi.org/10.1111/j.1540-4781.1988.tb04178.x

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology: An International Review*, *54*(2), 199- 231. https://psycnet.apa.org/doi/10.1111/j.1464-0597.2005.00205.x

Boekaerts, M., Pintrich, P. R., & Zeidner, M. (Eds.). (2000). *Handbook of self-regulation*. Academic Press.

Chamot, A. U. (2014, December). Developing self-regulated learning in the language classroom. In W. M. Chan, K. N. Chin, & T. Suthiwan (Eds.), *Proceedings of CLaSIC 2014* (pp. 78-88). Singapore University. http://www.fas.nus.edu.sg/cls/CLaSIC/clasic2014/Proceedings/chamot_annauhl.pdf

DuBay, W. H. (Ed.) (2006). *Classical readability studies*. Impact Information.

Gersten, R., Fuchs, L., Williams, J., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of research. *Review of Educational Research*, *71*(2), 279-320. https://doi.org/10.3102/00346543071002279

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge University Press.

Grabe, W. (2014). Key issues in L2 reading development. *Proceedings of the 4th CELC Symposium for English Language Teachers-Selected Papers* (pp. 8-18). National University of Singapore. http://www.nus.edu.sg/celc/research/books/ 4th%20 Symposium%20proceedings/2).%20William%20Grabe.pdf

Graham, S., & Harris, K. R. (2007). Best practices in teaching planning. In S. Graham, C. A. MacArthur, & J. Fitzgerald (Eds.), *Best practices in writing instruction* (pp. 119-140). Guilford Press.

Hamoulah Mardani, N., & Afghari, A. (2017). Self-regulated strategy development (SRSD) and the reading process: Effects on reading and metacognitive awareness. *Journal of Applied Linguistics and Language Research, 4*(6), 192-200. http://www.jallr.com/index.php/JALLR/article/view/673

Harris, K. R. , & Graham, S. (1996). *Making the writing process work: Strategies for composition and self-regulation*. MABrookline.

Harris, K. R., & Graham, S. (1996). *Making the writing process work: Strategies for composition and self-regulation.* Brookline

Harris, K. R., & Graham, S. (1999). Programmatic intervention research: Illustrations from the evolution of self-regulated strategy development. *Learning Disability Quarterly, 22*, 251-262.

Hashemi, L., & Thomas, B. (2009). *Objective PET for schools (Practice Test Booklet with Answers)*. Cambridge University Press.

Hedin, L., Mason, L. H., & Gaffney, J. (2011). TWA plus prompted discourse: Reading comprehension effects for three students with learning disabilities. *Preventing School Failure, 55*, 148-157.

Housand, A., & Reis, S. M. (2008). Self-regulated learning in reading: Gifted pedagogy and instructional settings. *Journal of Advanced Academics, 20*(1), 108-136. https://doi.org/10.4219/jaa-2008-865

Mason, L. H. (2004). Explicit self-regulated strategy development versus reciprocal questioning: Effects on expository reading comprehension among struggling readers. *Journal of Educational Psychology, 96*(20), 283-296. https://doi.org/10.1037/0022-0663.96.2.283

Mason, L. H. (2013). Teaching students who struggle with learning to think before, while, and after reading: Effects of self-regulated strategy development instruction. *Reading & Writing Quarterly, 29*(2), 124-144. https://psycnet.apa.org/doi/10.1080/10573569.2013.758561

Mason, L. H., Dunn Davison, M., Hammer, C. S., Miller, C. A., & Glutting, J. (2012). Knowledge, writing, and language outcomes for a reading comprehension and writing intervention. *Reading and Writing: An Interdisciplinary Journal*, *26*(7), 1133-1158.

Mason, L. H., Meadan-Kaplansky, H., Hedin, L., & Corso, L. (2006). Self-regulated strategy development instruction for expository text comprehension. *Teaching Exceptional Children*, *38*, 47-52. https://doi.org/10.1177/004005990603800407

Mason, L. H., Meadan-Kaplansky, H., Hedin, L., & Taft, R. (2013). Self-regulating informational text reading comprehension: Perceptions of low-achieving students. *Exceptionality: A Special Education Journal, 21*(2), 69-86. https://doi.org/10.1080/09362835.2012.747180

Mbato, C. L. (2013). *Facilitating EFL learners' self-regulation in reading: Implementing a metacognitive approach in an Indonesian higher education context* [Master's thesis, Southern Cross University].

Moos, D. C., & Ringdal, A. (2012). Self-regulated learning in the classroom: A literature review on the teacher's role. *Education Research International*, *1*(15), 1-16. https://doi.org/10.1155/2012/423284

Mori, S. (2004). Significant motivational predictors of the amount reading by EFL learners in

Japan. *RELC, 35*(1), 63-81. https://doi.org/10.1177/003368820403500106

Morshedian, M., Hemmat, F., & Sotoudehnama, E. (2017).Training EFL learners in self-regulation of reading: Implementing an SRL model. *Reading & Writing Quarterly, 33*(3), 290-303. https://doi.org/10.1080/10573569.2016.1213147

Nation, I. S. P. (2009). *Teaching ESL/EFL reading and writing*. Routledge.

Pallant. J. (2005). *SPSS survival manual: A step by step guide to data analysis using SPSS for windows* (Version 12) (2nd ed.). Allen & Unwin.

Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist, 36*(2), 89-101. https://psycnet.apa.org/doi/10.1207/S15326985EP3602_4

Perry, N. E. (1998). Young children's self-regulated learning and contexts that support it. *Journal of Educational Psychology, 90*(4), 715-729. https://psycnet.apa.org/doi/10.1037/0022-0663.90.4.715

Pintrich, P. R. (2000). The goal of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 452-502). Academic Press.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement, 53*(3), 801-813. https://doi.org/10.1177/0013164493053003024

Pintrich, P. R., & Zusho, A. (2007). Student motivation and self-regulated learning in the college classroom. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 731-810). Springer.

Roohani, A., & Asiabani, S. (2015). Effects of self-reading comprehension and metacognition. *GEMA Online Journal of Language Studies, 15*(3), 31-49.

Roohani, A., Hashemian, M., Asiabani, S. (2016). The effectiveness of strategic instruction on metacognition of EFL learners. *International Journal of Language Learning and Applied Linguistics World (IJLLALW), 11*(1), 1-9.

Roohani, A., Hashemian, M., Kazemian, Z. (2017). Effects of rhetorical analysis and self-regulation strategies on Iranian EFL learners' critical thinking and reading comprehension of argumentative texts. *Journal of Language Horizons, 1*(1), 53-75. https://doi.org/10.22051/lghor.2017.8769.1017

Souvignier, E., & Mokhlesgerami, J. (2006). Using self-regulation as a framework for implementing strategy instruction to foster reading comprehension. *Learning and Instruction*, *16*(1), 57-71.

https://psycnet.apa.org/doi/10.1016/j.learninstruc.2005.12.006

Swalander, L., & Taube, K. (2007). Influences of family-based prerequisites, reading attitude, and self-regulation on reading ability. *Contemporary Educational Psychology, 32*(2), 206-230. https://psycnet.apa.org/doi/10.1016/j.cedpsych.2006.01.002

Taylor, L, & Weir, C. J. (2012, May 29- June 3). *Automated approaches to establishing context validity in reading tests* [Conference session]. The 9th EALTA Conference, Innsbruck: Austria. https://core.ac.uk/download/pdf/286049534.pdf

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In P. Pintrich, M. Boekaerts, & M. Seidner (Eds.), *Handbook of self-regulation* (pp. 531-566). Academic Press.

Woolley, G. (2011). *Reading comprehension: Assisting children with learning difficulties*. Springer Science +Business Media B.V.

Zimmerman, B. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice, 41*(2), 64-72. http://dx.doi.org/10.1207/s15430421tip4102_2

Zimmerman, B. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*(1), 166-83. https://doi.org/10.3102/0002831207312909

Zwier, L. (2016). *Inside reading: Student book level 2* (2nd ed.). Oxford University Press.

Appendix A

MSLQ (from Pintrich et al., 1993)

The following questions ask about your motivation for and attitudes about this class. Remember there is no right or wrong answer; just answer as accurately as possible. Use the scale below to answer the questions. If you think the statement is very true of you, circle 7; if a statement is not at all true of you, circle 1. If the statement is more or less true of you, find the number between 1 and 7 that best describes you.

1	2	3	4	5	6	7
Not at all true of me	Not very much like me	Somewhat not like me	Somewhat like me	Like me	Much like me	Very true of me

Part A. Motivation

1	In a class like this, I prefer reading material that really challenges me so I can learn new things.	1	2	3	4	5	6	7
2	If I read in appropriate ways, then I will be able to understand the reading material in this course.	1	2	3	4	5	6	7
3	When I take a test I think about how poorly I am doing compared with other students.	1	2	3	4	5	6	7
4	I think I will be able to use what I learn in this course in other courses.	1	2	3	4	5	6	7
5	I believe I will receive an excellent grade in this class.	1	2	3	4	5	6	7
6	I'm certain I can understand the most difficult reading material presented in this course.	1	2	3	4	5	6	7
7	Getting a good grade in this class is the most satisfying thing for me right now.	1	2	3	4	5	6	7
8	When I take a test I think about items on other parts of the test I can't answer.	1	2	3	4	5	6	7
9	It is my own fault if I don't understand the reading material in this course.	1	2	3	4	5	6	7
10	It is important for me to understand the reading material in this class.	1	2	3	4	5	6	7
11	The most important thing for me right now is improving my overall grade point average, so my main concern in this class is getting a good grade	1	2	3	4	5	6	7
12	I'm confident I can learn the basic concepts taught in	1	2	3	4	5	6	7

	this course.							
13	If I can, I want to get better grades in this class than most of the other students.	1	2	3	4	5	6	7
14	When I take tests I think of the consequences of failing.	1	2	3	4	5	6	7
15	I'm confident I can understand the most complex reading material presented by the instructor in this course.	1	2	3	4	5	6	7
16	In a class like this, I prefer reading material that arouses my curiosity, even if it is difficult to learn.	1	2	3	4	5	6	7
17	I am very interested in the content areas of reading materials in this course.	1	2	3	4	5	6	7
18	If I try hard enough, then I will understand the reading material.	1	2	3	4	5	6	7
19	I have an uneasy, upset feeling when I take an exam.	1	2	3	4	5	6	7
20	I'm confident I can do an excellent job on the assignments and tests in this course.	1	2	3	4	5	6	7
21	I expect to do well in this class.	1	2	3	4	5	6	7
22	The most satisfying thing for me in this course is trying to understand the content and reading selections as thoroughly as possible.	1	2	3	4	5	6	7
23	I think the reading material in this class is useful for me to learn.	1	2	3	4	5	6	7
24	When I have the opportunity in this class, I choose reading assignments that I can learn from even if they don't guarantee a good grade.	1	2	3	4	5	6	7
25	If I don't understand the reading material, it is because I didn't try hard enough.	1	2	3	4	5	6	7
26	I like the reading selections of this course.	1	2	3	4	5	6	7
27	Understanding the reading selections of this course is very important to me.	1	2	3	4	5	6	7
28	I feel my heart beating fast when I take an exam	1	2	3	4	5	6	7
29	I'm certain I can master the reading skills being taught in this class.	1	2	3	4	5	6	7
30	I want to do well in this class because it is important to show my ability to my family, friends, employer, or others.	1	2	3	4	5	6	7

31	Considering the difficulty of this course and its reading selections, the teacher, and my skills, I think I will do well in this class.	1	2	3	4	5	6	7
Part B. Learning Strategies								
32	When I study the readings for this course, I outline the material to help me organize my thoughts.	1	2	3	4	5	6	7
33	During class time I often miss important points because I'm thinking of other things.	1	2	3	4	5	6	7
34	When reading in this course, I often try to explain the reading material to a classmate or friend.	1	2	3	4	5	6	7
35	I usually read in a place where I can concentrate on my reading work.	1	2	3	4	5	6	7
36	When reading in this course, I make up questions to help focus my reading.	1	2	3	4	5	6	7
37	I often feel so lazy or bored when I read in this class that I quit before I finish what I planned to do.	1	2	3	4	5	6	7
38	I often find myself questioning things I read in this course to decide if I find them convincing.	1	2	3	4	5	6	7
39	When I read in this class, I practice saying the material to myself over and over.	1	2	3	4	5	6	7
40	Even if I have trouble reading the material in this class, I try to do the work on my own, without help from anyone.	1	2	3	4	5	6	7
41	When I become confused about something I'm reading in this class, I go back and try to figure it out.	1	2	3	4	5	6	7
42	When I read in this course, I go through the readings and my class notes and try to find the most important ideas.	1	2	3	4	5	6	7
43	I make good use of my reading time for this course.	1	2	3	4	5	6	7
44	If course readings are difficult to understand, I change the way I read the material.	1	2	3	4	5	6	7
45	I try to work with other students from this class to complete the reading assignments.	1	2	3	4	5	6	7
46	When reading in this course, I read my course readings over and over again.	1	2	3	4	5	6	7
47	When a theory, interpretation, or conclusion is presented	1	2	3	4	5	6	7

	in the reading selections, I try to decide if there is good supporting evidence.							
48	I work hard to do well in this class even if I don't like what we are doing.	1	2	3	4	5	6	7
49	I make simple charts, diagrams, or tables to help me organize reading material.	1	2	3	4	5	6	7
50	When studying for this course, I often set aside time to discuss reading material with a group of students from the class.	1	2	3	4	5	6	7
51	I treat the reading material as a starting point and try to develop my own ideas about it.	1	2	3	4	5	6	7
52	I find it hard to stick to a reading schedule.	1	2	3	4	5	6	7
53	When I study for this class, I pull together information from different sources, such as lectures, readings, and discussions.	1	2	3	4	5	6	7
54	Before I read new reading material thoroughly, I often skim it to see how it is organized.	1	2	3	4	5	6	7
55	I ask myself questions to make sure I understand the material I have been reading.	1	2	3	4	5	6	7
56	I try to change the way I read in order to fit the course requirements and the instructor's teaching style.	1	2	3	4	5	6	7
57	I often find that I have been reading but don't know what it was all about.	1	2	3	4	5	6	7
58	I ask the instructor to clarify concepts in reading selections I don't understand well.	1	2	3	4	5	6	7
59	I memorize key words to remind me of important concepts in reading selections.	1	2	3	4	5	6	7
60	When a reading selection is difficult, I either give up or only read the easy parts.	1	2	3	4	5	6	7
61	I try to think through a topic and decide what I am supposed to learn from it rather than just reading it over when reading in this course.	1	2	3	4	5	6	7
62	I try to relate ideas in the reading selections in this course to the material in others whenever possible.	1	2	3	4	5	6	7
63	When I read in this course, I go over my class notes and make an outline of important concepts.	1	2	3	4	5	6	7

64	When reading for this class, I try to relate the material to what I already know.	1	2	3	4	5	6	7
65	I have a regular place set aside for reading.	1	2	3	4	5	6	7
66	I try to play around with ideas of my own related to what I am reading in this course.	1	2	3	4	5	6	7
67	When I study for this course, I write brief summaries of the main ideas from the readings and my class notes.	1	2	3	4	5	6	7
68	When I can't understand the reading material, I ask another student in this class for help.	1	2	3	4	5	6	7
69	I try to understand the material in this class by making connections between the readings and the concepts from the lectures.	1	2	3	4	5	6	7
70	I make sure that I keep up with the weekly reading assignments for this course.	1	2	3	4	5	6	7
71	Whenever I read an assertion or conclusion in this class, I think about possible alternatives.	1	2	3	4	5	6	7
72	I make lists of important items in this course and memorize the lists.	1	2	3	4	5	6	7
73	I attend this class regularly.	1	2	3	4	5	6	7
74	Even when reading materials are dull and uninteresting, I manage to keep working until I finish.	1	2	3	4	5	6	7
75	I try to identify students in this class whom I can ask for help if necessary.	1	2	3	4	5	6	7
76	When reading in this course I try to determine which reading parts I don't understand well.	1	2	3	4	5	6	7
77	I often find that I don't spend very much time on reading the material in this course because of other activities.	1	2	3	4	5	6	7
78	When I read in this class, I set goals for myself in order to direct my activities in each reading period.	1	2	3	4	5	6	7
79	If I get confused taking notes when reading, I make sure I sort it out afterwards.	1	2	3	4	5	6	7
80	I rarely find time to review my notes or readings before an exam.	1	2	3	4	5	6	7
81	I try to apply ideas from reading selections to activities in other class such as lecture and discussion.	1	2	3	4	5	6	7

Journal of Language Horizons

Scientefic Quarterly Journal of Language Horizons, *Alzahra University*
Volume 7, Issue 1, Spring 2023 (Serial No. 15)

The Impact of Test Length on Raters' Mental Processes During Scoring Test-Takers' Writing Performance

Research Article
pp. 159-182

Fateme Nikmard[1]
Kobra Tavassoli*[2]

Received: 2021/08/19 Accepted: 2022/02/12

Abstract

Different factors such as the writing genre, writing prompt, and/or test length can influence the raters' mental processes while scoring writing tests. Accordingly, whether an increase or a decrease in test length has any impact on how raters evaluate test-takers' writing performance was the motive underlying this research. For this purpose, 12 EFL students who scored between 5.5 to 7.5 on the writing section of a mock IELTS test were selected based on availability sampling. The participants wrote three argumentative essays (the original, longer, and shorter versions). The three versions from each test-taker were then scored by three raters using IELTS task 2 writing band descriptors. Meanwhile, the raters provided verbal protocols explaining in detail the reasons underlying their scores to each test-taker's essay. Then, the verbal protocols were transcribed and content analyzed using Nvivo version 11 to extract the themes mentioned by the raters in scoring each writing test. The results showed that the raters paid more attention to certain factors in the band descriptors and ignored some other factors. However, there was a similar pattern among the raters in scoring the three writing tests. The results did not show any significant differences in the raters' mental processes while scoring each of the three writing tests. The conclusion was that test length is not a determining factor

* Corresponding Author
[1] PhD Candidate, Department of ELT, Faculty of Literature and Foreign Languages, Karaj Branch, Islamic Azad University, Karaj, Iran. fateme.nikmard@kiau.ac.ir
[2] Assistant Professor, Department of ELT, Faculty of Literature and Foreign Languages, Karaj Branch, Islamic Azad University, Karaj, Iran. kobra.tavassoli@kiau.ac.ir

DOI: 10.22051/lghor.2022.37340.1545
DOR: 20.1001.1.2588350.2023.7.1.7.5

influencing the mental processes of raters in writing tests. Therefore, raters and test developers do not need to worry about test length influencing the raters' scoring.

Keywords: mental processes, rater, test length, verbal protocol, writing test

Introduction

Evaluation is an integral part of almost all learning processes and it is defined as the systematic gathering of information to make appropriate decisions (Alderson & Banerjee, 2002). Appropriate decision-making and the results drawn from it can greatly influence the test-takers' lives. Thus, people involved in decision-making processes, such as raters, are considered important in any act of evaluation, especially in evaluating test-takers' performance on speaking and writing tests as the scores may depend on their subjective viewpoints. Of course, there are different solutions to overcome the problem of subjectivity, one of the most important of which is to have more than one rater in evaluating test-takers' performance (Alderson & Banerjee, 2002). However, rater subjectivity and bias may be influenced by various factors such as the length of the tests, and it is vital to do research on such factors to reduce subjectivity and bias as much as possible. For instance, a longer writing test might give raters the impression that the writer is well-informed about a subject and lead them to assign a different score to such a test, while a shorter test might convey the opposite impression. In fact, Ackerman and Kanfer (2009) believed that an important factor that may influence the raters' scoring, especially in writing tests, is test length since the examinees' performance may change when the number of words they should write differs. Another important factor influencing the raters' scoring is their characteristics, such as their mental processing, which is considered a prominent factor in Eckes's (2012) perspective.

A close look at the related literature shows that even though the factors influencing rater subjectivity and bias in assigning scores to the examinees' performance have been the subject of a plethora of studies (e.g., Barkaoui, 2019; Humphrey-Murto et al., 2021; Weigle, 2002; Wind, 2019), there is a gap in the literature in investigating the effect of test length on the raters' mental processes in assigning scores to writing tests and the difference it may cause in their assigned scores to the best of the researchers' knowledge. However, a similar study was done by Hoora (2019) on investigating the examiners' mental processes when evaluating

three versions of the IELTS speaking test with different lengths. Thus, the present study aimed to check the impact of test length on raters' evaluation of test-takers' writing performance through identifying their mental processes when scoring three similar essays. Test length in this study refers to the number of words test-takers were required to write when participating in the writing tests. We asked the participants to write argumentative essays on three topics and named them the original, longer, and shorter versions. The results of such a study can be useful for test designers since if it shows that raters' mental processes differ while they score different versions of a writing test, it is necessary to think about test length as a determining factor influencing the test-takers' scores and the decisions made on their future lives.

Literature Review

Writing and its Evaluation

Writing is a common activity in which writers construct the required response and it is frequently used in most of the standardized writing assessments such as GRE, TOEFL iBT, and IELTS, to name a few (Eckes, 2012). However, as Eckes stated, a piece of writing is not a product being produced by an individual in isolation, rather, writing is assumed to be socially and culturally produced. A piece of writing is created in a particular context, is aimed to achieve a specific goal, and is directed towards a special class of audience. To produce an acceptable piece of writing, which is an extremely complex task, writers need to have a good command of words, know how to write grammatically correct sentences, be familiar with the organization of paragraphs, be familiar with the mechanics of writing, pay attention to cohesion and coherence, and develop the content appropriately (Weigle, 2002). Sahragard and Mallahi (2014) further added that when writing, learners should set a goal for it, plan their writing carefully, think about its format and structure, and finally revise it carefully. According to Cohen (2003), how a test-taker does the task of writing is an essential factor in the successful use of the cognitive and metacognitive processes associated with the writing task. Writing in a second/foreign language is even more demanding considering any of these aspects, therefore, it is essential to teach these aspects to students, as well as to pay attention to them in evaluating their writing to ensure students have learned the essential

points successfully.

A crucial aspect of writing is its evaluation (Eckes, 2012). In evaluating writing, one of the three scales of primary trait, holistic, or analytic can be used. The analytic scoring scale contains different writing components, such as organization, register, coherence, cohesion, mechanics of writing, content, and accuracy of linguistic devices. The good point about this kind of rating scale is that each element can be marked independently of the other components (Sawaki, 2007). Moreover, it has more benefits over the other scales. One of its major advantages is that since it provides information on different aspects of a piece of writing, it is more useful to recognize the writer's specific writing abilities and weaknesses (Eckes, 2012). The analytic scale can also be used as a criterion when training raters since it breaks writing into different subcategories such as vocabulary, grammar, content, organization, and mechanics of writing, by assigning a separate score to each. In other words, each writing should be scored based on the different aspects of the scale (Brown, 2005). Moreover, since different aspects of a test-taker's ability are evaluated separately, it becomes clear in which domain(s) a test-taker is strong or weak. The last important advantage of analytic scoring is the high inter-rater reliability it creates among the raters (Weigle, 2002). Because of these advantages, analytic scoring was used in this study, too.

It is worth mentioning that because of its importance, writing has been the subject of an abundance of studies throughout the years. Different aspects of writing have already been investigated, for example, the writing resources second language learners use in their writing (Oh, 2020), the effect of planning time and different task conditions on EFL learners' writing complexity, accuracy, and fluency (Fazilatfar et al., 2020), the sources of inconsistency in L2 writing scores (Barkaoui, 2019), the degree of similarities and differences between the students' performance on writing tasks and actual academic writing tasks (Llosa & Malone, 2019), and the effect of task-repetition and elicitation on the learners' writing (Asadi Vahdat & Tavassoli, 2019), to name some. However, the raters' mental processes (i.e., the way they interpret and use the rating scale) when evaluating test-takers' performance in writings of different lengths have rarely been investigated.

Raters

There are many issues involved in the process of interpreting the results of language assessment, one of which is the issue of raters in speaking and writing tests, especially when there is more than one rater scoring the same task produced by the same test-takers (Marcoulides & Ing, 2014). Raters have an influential role in evaluating writing and speaking tests since their knowledge and experience might affect the scores they assign (Duijm et al., 2018). Additionally, the raters' background seems to be a factor determining the language features they pay attention to in a piece of writing (Bijani, 2018). Kang and Veitch (2017) also confirmed the effect of raters' backgrounds on the way they assessed ESL writing pieces. In addition, rater training is considered as another influential factor (Bijani, 2018). Further, Attali (2016) investigated rater training and found that training had a parallel effect to that of experience. In other words, the scores the newly-trained raters assigned were close to those of experienced raters. On the other hand, Duijm et al. (2018) tried to minimize background effects by training raters and giving them detailed instructions, but raters in their study tended to assign dissimilar scores to the same performance due to their different backgrounds. Wind (2019) also believed that the differences in the judgments provided by various raters are mainly due to their construct-irrelevant characteristics, and such differences can threaten the fairness of their scores. Thus, researching different characteristics of raters, such as their background, preceding experience, mother tongue, and tolerance for errors like word order and verb form (Huang & Foote, 2010) can be quite informative about how they score writing and speaking tests, and how the consistency of their scoring can be improved.

One more determining factor in the scores raters assign to the test-takers' oral and written performance is the rating scale used (Purpura, 2004). Purpura believed that the analytic rating scale is a better choice in comparison to other scales since it provides the raters with detailed explanations on each important component to be measured. However, raters may translate the explanations differently and assign different scores to the same performance which may lead to bias and unreliable test scores. Purpura (2004) also offered some techniques to reduce the effects of such matters. He recommended using as clear and detailed scales as possible, training raters regarding the rubric, utilizing sample performances in

training sessions so that raters get familiar with different points of the scale, asking a third rater to judge the scores in case there is a huge inconsistency in the scores assigned by the first two raters, and continuously observing the raters and providing them with feedback if necessary.

Variations in ratings can be attributed to differences in the raters' mental processes, too (Esfandiari & Noor, 2019). That is, the scores assigned to the examinees' performance are not merely the result of the quality of their performance. The rater's cognition is also involved since the scores they assign are affected by their ability to compare and contrast the mental representations of the examinees' responses with their own mental processes (Purpura, 2014). Raters' cognitive processes, which are related to the structure of the human information processing system, can affect the way raters assign scores and the strategies they use in the act of rating (Han, 2016).

One common technique to come up with the main cognitive and/or metacognitive processes involved in the raters' scoring is using and analyzing their verbal protocols (May, 2011) through which it is possible to find out the features often emphasized by the raters (Ducasse, 2010). Lumley (2005) introduced verbal protocol (also known as think-aloud) as a way of data gathering in which the participants are required to either think aloud as they are performing a task (called introspective) or after they finished it (called retrospective). In verbal protocols, participants are asked to pronounce what comes to their minds regarding their performance. Such information is useful since it provides the researcher with good insights into the participants' cognitive processes. Formal verbal protocols are first recorded, then transcribed, and finally analyzed (Lumley, 2005).

In recent years, the raters' cognitive processes are being studied by different researchers. For instance, Esfandiari and Noor (2019) utilized a 4-stage processing model proposed by Han (2016) to investigate the cognitive processes two groups of raters (novice and expert) followed when they rated the examinees' responses to a speaking task. They concluded that different degrees of expertise have significant effects on the decisions made on different aspects of responses since they interpreted the responses differently and paid attention to different aspects of the criteria to judge the responses.

In another study, Humphrey-Murto et al. (2021) investigated the effect the

prior familiarity of the raters with the examinees has on the raters' judgment. They found that familiarity makes the raters somehow biased either towards or away from the examinees; the raters' negative prior evaluation of the examinees had a greater negative influence. They also found that, although the raters' expertise or the training they receive cannot reduce such effects, their higher levels of accountability, following particular standards, and also decreasing their cognitive role can play a reducing role for such effects.

Test Length

Test length as one of the aspects of the expected response, itself a major test method facet (Bachman, 1990), can be an influential factor influencing not only the test-takers' performance but also the raters' scoring of that performance. In fact, test method facets, including the rubric, environment, input, expected response, and the relationship between input and expected response, are probable factors underlying the inconsistencies in test-takers' performance, raters' mental processes while scoring, test interpretations, and inferences made about test-takers' abilities (Weigle, 2002). The length of a response, which can vary from a word to an essay, may be effective on the test-takers' performance since the longer the output is, the more is the possibility of the effect of intervening factors such as the knowledge of vocabulary and grammar.

Furthermore, the length of a test seriously influences productivity measures whose aim is to assess lexical or grammatical diversity (Shirai & Vercellotti, 2014). Lexical diversity, for instance, can be measured by the ratio of type to token. That is, lexical diversity is usually calculated by dividing the number of various words used in a text (type) over the entire number of used words (token). In other words, longer texts tend to achieve higher scores due to the more space they provide for the writers to use more words of different types. On the other hand, to measure grammatical complexity, the number of dependent clauses is counted in each T-unit and then averaged across all the T-units used throughout the text (Biber et al., 2011). The assumption is that using more subordinate clauses is a sign of more grammatical complexity.

Regarding the writing ability of language learners or test-takers, length is considered as an indirect measure of development; that is, it may help the writer to

elaborate more on a topic with fluidity and therefore gain a higher score. Such flexibility is then one of the common indications of the writer's proficiency level. In other words, a writer with a higher proficiency level is able to respond longer and a writer with a lower proficiency level produces a shorter response (Plakans, 2014).

Test length in different types of tests was sporadically investigated before. Ackerman and Kanfer (2009) researched the relationship between test length and cognitive fatigue. Although the more time spent on the test caused a kind of fatigue for the test-takers, it improved the quality of their performance. In a more recent study, Sahin and Anil (2017) used three unidimensional dichotomous models of item response theory (IRT) to discover the possible impacts of the two factors of test length and sample size on item parameters. They concluded that the synthesis of the two factors of test length and the sample size is more prominent than each individual factor.

Thus, it seems important to do more research on test length to examine how it might influence the raters' mental processes in scoring test-takers' writing ability. The present research was an attempt in this regard. Accordingly, the following research question was posed:

Does an increase or a decrease in test length make any difference in the raters' mental processes when they score test-takers' writing performance?

Method

This study was carried out through quantitative content analysis, in which themes and subthemes were extracted from the collected data (Dörnyei, 2007). According to Coe and Scacco (2017), content analysis presents descriptions of the data, and then, it is possible to make generalizations about the available patterns within the data. This is the quantification of patterns or coding where there are instructions about the features to be derived from the data.

Participants

Two groups of participants took part in the present study. The first group consisted of 12 female EFL students with the age range of 28-36 years old who were selected based on availability sampling. In order to ensure the homogeneity of these participants, only those who scored between 5.5-7.5 based on IELTS band

descriptors of writing were selected.

The second group of participants consisted of three female raters who were experienced EFL teachers and raters using IELTS band descriptors in IELTS preparation classes. Their age range was 31-37 and their teaching and rating experience ranged from 7 to 13 years. We tried to control the raters' gender, age, and experience to minimize their potential effect on the findings of the study. In addition, to ensure consistency in scoring, the raters participated in a one-hour training session to provide them with instructions on how to use the scale in this study and how to provide verbal protocols on their rating.

Tables 1 and 2 show the demographic information of the two groups of participants. Both groups participated willingly in this study.

Table 1

Demographic Information of the Test-Takers

Student	Age	Gender	IELTS Score
1	28	Female	6.5
2	32	Female	6
3	35	Female	7.5
4	29	Female	6.5
5	31	Female	6.5
6	35	Female	6
7	36	Female	6.5
8	30	Female	5.5
9	35	Female	6
10	32	Female	5.5
11	29	Female	6.5
12	34	Female	6.5

Table 2

Demographic Information of the Raters

Rater	Age	Gender	Experience
1	37	Female	13 years
2	31	Female	7 years
3	35	Female	11 years

Materials and Instruments

We asked the EFL students to write three argumentative essays taken from the Cambridge English IELTS 10 (2015) and named the three writings as the original version, the longer version, and the shorter version.

In the original version of the writing test, the test-takers were required to write 250 words in 40 minutes, which is the case in the IELTS writing exam. On the other hand, in the longer version of the writing test, they were asked to write 300 words in 50 minutes, whereas in the shorter version, the word limit was reduced to 200 words and time was limited to 30 minutes. The time limit and word count in the longer and shorter versions of the writing test were calculated mathematically by dividing the word count by the time limit in the original version and rounding up the numbers. Such time allotment was also used in Ahmad's (2021) study in which he allocated 30 and 50 minutes as low-timing and long-timing conditions for test-takers to write two IELTS argumentative essays.

The topics for the three writing tests were as the following.

- *The original version*:

You should spend about 40 minutes on this task.
Write about the following topic. Write at least 250 words.

It is important for children to learn the differences between right and wrong at an early age. Punishment is necessary to help them learn this distinction. To what extent do you agree or disagree with this opinion? What sort of punishment should parents and teachers be allowed to use to teach good behavior to children?

- *The longer version*:

You should spend about 50 minutes on this task.
Write about the following topic. Write at least 300 words.

Some people think that all university students should study whatever they like. Others believe that they should only be allowed to study subjects that will be useful in the future, such as those related to science and technology. Discuss both these views and give your own opinion.

- *The shorter version*:

You should spend about 30 minutes on this task.
Write about the following topic. Write at least 200 words.

Every year several languages die out. Some people think that this is not

important because life will be easier if there are fewer languages in the world. To what extent do you agree or disagree with this opinion?

In addition, the IELTS task 2 writing band descriptors was used as the scoring rubric to score the three essays in this study. The scale consists of four main components of task achievement, coherence and cohesion, lexical resource, and grammatical range and accuracy, each of which is composed of detailed descriptions on how to be evaluated.

Also, the three raters were asked to provide a verbal protocol/report when scoring each of the three versions of the participants' writings by explaining in detail the reasons underlying their scores. These verbal protocols were recorded and later transcribed and content analyzed through the Nvivo software version 11.

Procedure

Twelve EFL female students were selected based on availability sampling and their willingness to participate in the study. Each test-taker was asked to write three argumentative essays with different time limits and word counts. At first, the original version of the writing test, derived from Cambridge English IELTS 10 (2015), was administered to the students. That is, they were asked to write an argumentative essay of 250 words in 40 minutes on a certain topic as it is the case in the IELTS exam. After two weeks, the participants were asked to write a longer argumentative essay of 300 words in 50 minutes on another topic. Finally, after another interval of two weeks, the participants wrote a shorter argumentative essay of 200 words in 30 minutes on another topic. The 12 participants were those who scored 5.5-7.5 based on IELTS task 2 writing band descriptors in the original version of the test. The reason for using argumentative topics in the three versions of the writing test was to keep the variation due to genre of writing to the minimum and the two-week interval was set to reduce the potential effect of genre on the test-takers' performance.

In the next phase of the study, the three raters, who passed a one-hour training session, scored the three writings of each test-taker based on the IELTS rubric and simultaneously provided a verbal protocol/report on how and why they scored each writing and what points in the scoring rubric they paid more attention to. The verbal protocols were then transcribed, saved in three separate files (the

original, longer, and shorter versions) in Nvivo version 11, and later content analyzed to answer the research question of the study.

Results

The Coding System

The IELTS band descriptors have four criteria including (1) task achievement, (2) coherence and cohesion, (3) lexical resource, and (4) grammatical range and accuracy, which were identified as the major themes in this study. In addition, some subthemes were identified for each of these themes based on the details in the IELTS band descriptors and the data extracted from the raters' verbal protocols. The subthemes other than those that existed in the IELTS scale, which were extracted from the verbal protocols, were also categorized under the four major themes of the IELTS scale. The process of identifying and classifying the subthemes was done by the researchers who worked collaboratively to do this. We cross-checked the subthemes and their classification under the four themes with another researcher familiar with this process and made the necessary modifications. The classification of themes and subthemes is represented in Table 3.

Table 3

The Themes and Subthemes Extracted from the IELTS Scoring Rubric and the Raters' Verbal Protocols

Task Achievement	Coherence and Cohesion	Lexical Resource	Grammatical Range and Accuracy
Addressing the task	Paragraphing	Range of vocabulary	Range of structures
Writer's position towards the topic	Sequencing information and ideas	Using less/uncommon lexical items	Error-free sentences
Extending and supporting ideas	Aspects of cohesion	Word features (e.g., spelling, word choice, word formation)	Using complex structures
Answering the question	Progression of ideas		Punctuation

Non-overgeneralization of ideas	Using cohesive devices	Avoiding grammatical errors
Conclusion	Clear presentation of the central topic	
Main ideas	Referencing	
Format		
Avoiding irrelevant details		

Investigation of the Research Question

At first, the transcriptions of all the verbal protocols were imported into three Nvivo files, one for each version of the writing test (the original, longer, and shorter versions). Then, the content of each file was analyzed word by word to find out the themes and subthemes each rater mentioned in scoring each writing along with their frequency of occurrence. Tables 4-7 present sample examples from the raters' verbal protocols for the subthemes of *Task Achievement*, *Coherence and Cohesion*, *Lexical Resource*, and *Grammatical Range and Accuracy*, respectively.

Table 4

Examples for the Subthemes of Task Achievement

Subthemes	Example
Addressing the task	She addressed all parts of the task.
Writer's position towards the topic	She presented a clear position about the topic.
Extending and supporting ideas	She presented, extended, and supported ideas well.
Answering the question	She presented well-developed responses to the questions with relevant extended and well-supported ideas.
Non-overgeneralization of ideas	There is a tendency to overgeneralize, and supporting ideas lack focus.
Conclusion	Although she presented the conclusion, it may be unclear or repetitive.
Main ideas	Some limited main ideas are presented.
Format	The format is inappropriate in some places of this writing.
Avoiding irrelevant details	There are some irrelevant details and ideas.

Table 5

Examples for the Subthemes of Coherence and Cohesion

Subthemes	Example
Paragraphing	She used paragraphing but not always logically.
Sequencing information and ideas	The writing was logically organized.
Aspects of cohesion	She managed all aspects of cohesion well.
Progression of ideas	There is a clear progression of ideas throughout the writing and the passage.
Using cohesive devices	She used a range of cohesive devices appropriately throughout the essay.
Clear presentation of the central topic	She presented a clear central topic within each paragraph.
Referencing	She did not always use referencing clearly and appropriately.

Table 6

Examples for the Subthemes of Lexical Resource

Subthemes	Example
Range of vocabulary	She used a sufficient range of vocabulary that allowed some flexibility and precision.
Using less/uncommon lexical items	She used less common lexical items with some awareness of style and collocation.
Word features	She made some errors in spelling but they do not impede communication.

Table 7

Examples for the Subthemes of Grammatical Range and Accuracy

Subthemes	Example
Range of structures	She used a wide range of structures.
Error-free sentences	The majority of the sentences were error-free, and she made only very occasional errors.
Using complex structures	She used a variety of complex structures.
Punctuation	She had good control of punctuation.
Avoiding grammatical errors	She avoided some grammatical errors.

Next, Tables 8-11 report the frequency of each subtheme of the four major themes in the original, longer, and shorter versions of the writing test, along with the significance values of chi-squares to check whether the differences in each subtheme and major theme along the three writing tests were significant or not.

Table 8

The Frequency of the Subthemes of Task Achievement

Task Achievement	Original Writing Test	Longer Writing Test	Shorter Writing Test	Significance Value of Chi-Square
Addressing the task	35	33	32	.93
Writer's position towards the topic	31	27	30	.86
Extending and supporting ideas	22	26	22	.79
Answering the question	1	3	3	.56
Non-overgeneralization of ideas	6	4	3	.58
Conclusion	11	7	8	.60
Main ideas	25	31	27	.71
Format	3	0	6	.31
Avoiding irrelevant details	0	3	2	.65
Total	134	134	133	.99

As shown in Table 8, *addressing the task*, *writer's position towards the topic*, and *main ideas* were the most referred subthemes in all the three versions of the test. This shows the importance of these three subthemes for the raters. On the other hand, *non-overgeneralization of ideas*, *answering the question*, *format*, and *avoiding irrelevant details* were the subthemes the raters mentioned the least again in the three versions of the test, which shows their lower significance for the raters. The other two subthemes under this major theme (i.e., *extending and supporting ideas*, and *conclusion*) fell between these two extremes. Looking at the significance

values of chi-square comparing the differences in frequencies of each subtheme across the three versions of the writing test, it was concluded that there was not a statistically significant difference in the frequencies of the subthemes of *Task Achievement* mentioned by the three raters in the three tests since all the significance values are higher than .05 probability level (α = .05; $p > \alpha$). Further, the total frequencies of the subthemes of *Task Achievement* were very close in the three versions of the test and the related significance value of chi-square was higher than the critical .05 level (p = .99; α = .05; $p > \alpha$). It means there was not a statistically significant difference in this major theme among the different versions of the test, either. Overall, the conclusion about the major theme of *Task Achievement* was that although there were subtle differences in the frequencies of its subthemes and the total frequency mentioned by the raters, test length did not make any considerable differences in the mental processes the raters mentioned regarding *Task Achievement* while scoring the three essays.

Table 9

The Frequency of the Subthemes of Coherence and Cohesion

Coherence and Cohesion	Original Writing Test	Longer Writing Test	Shorter Writing Test	Significance Value of Chi-Square
Paragraphing	30	28	31	.92
Sequencing information and ideas	28	26	29	.91
Aspects of cohesion	5	8	7	.70
Progression of ideas	20	17	22	.72
Using cohesive devices	29	29	28	.98
Clear presentation of the central topic	11	10	5	.30
Referencing	3	2	7	.17
Total	126	120	129	.84

Based on the information in Table 9, *paragraphing* was the most frequent and therefore the most important subtheme in the raters' ideas in the three versions

of the test, which was followed by *using cohesive devices*, *sequencing information and ideas*, and *progression of ideas*, whereas *referencing* was the least mentioned subtheme, followed by *aspects of cohesion*, and *clear presentation of the central topic*. Here again, the significance values of chi-square were all above the critical value of .05 ($\alpha = .05$; $p > \alpha$), meaning that there were no significant differences in the frequencies of each subtheme of *Coherence and Cohesion*. Once more, the total frequencies of the subthemes of *Coherence and Cohesion* were close to each other and the related significance value of chi-square was non-significant and above the critical value of .05 ($p = .84$; $\alpha = .05$; $p > \alpha$). Thus, there was not any substantial difference between the raters' viewpoints regarding the importance of the subthemes and the major theme of *Coherence and Cohesion* in the three versions of the test.

Table 10

The Frequency of the Subthemes of Lexical Resource

Lexical Resource	Original Writing Test	Longer Writing Test	Shorter Writing Test	Significance Value of Chi-Square
Range of vocabulary	33	33	32	.99
Using less/uncommon lexical items	18	16	13	.66
Word features	32	30	33	.92
Total	83	79	78	.91

As it can be seen in Table 10, *range of vocabulary* and *word features* (e.g., spelling, word choice, and word formation) were the most frequently mentioned subthemes of *Lexical Resource* by the raters in the three versions of the test, whereas the subtheme of *using less/uncommon lexical items* was the least frequently mentioned one, which shows its less significance in the raters' views. Once again, the results of the significance values of chi-square showed that there were no significant differences in the frequency of the subthemes of *Lexical Resource* as the related significance values of chi-square were all above the critical value of .05 ($\alpha = .05$; $p > \alpha$). In addition, the total frequencies of the subthemes of *Lexical Resource* were not much different from each other in the three versions of the test and the

related significance value of chi-square showed a non-significant difference since it was above the critical value of .05 (p = .91; α = .05; $p > \alpha$). To summarize, regarding the subthemes as well as the major theme of *Lexical Resource*, there was no significant difference between the raters' ideas in scoring the three versions of the writing test.

Table 11

The Frequency of the Subthemes of Grammatical Range and Accuracy

Grammatical Range and Accuracy	Original Writing Test	Longer Writing Test	Shorter Writing Test	Significance Value of Chi-Square
Range of structures	26	24	32	.56
Error-free sentences	10	10	5	.36
Using complex structures	13	21	16	.37
Punctuation	14	24	23	.22
Avoiding grammatical errors	28	29	25	.85
Total	91	108	101	.48

Finally, the information in Table 11 shows that *avoiding grammatical errors*, *range of structures*, and *punctuation* were the most-frequently mentioned subthemes of *Grammatical Range and Accuracy* by the raters, and therefore, the most important ones in the three versions of the test from the raters' perspectives. On the other hand, *error-free sentences* was the least mentioned subtheme by the raters, and therefore, the least important one again in the three versions of the test. The other subtheme under this major theme (i.e., *using complex structures*) fell between these two extremes. The same as the previous major themes, the significance values of chi-square for all the subthemes of *Grammatical Range and Accuracy* were above the critical .05 value (α = .05; $p > \alpha$), meaning that there were no statistically significant differences in the frequency of the mentioned subthemes. Further, the total frequencies of the subthemes of *Grammatical Range and Accuracy* were compared with each other. Here again, the related significance value for chi-square was non-significant and above the critical value of .05 (p = .48; α = .05; $p >$

α). Once more, the conclusion was that regarding the major theme of *Grammatical Range and Accuracy*, test length did not make any considerable differences in the mental processes the raters mentioned in scoring the three versions of the writing test.

To wrap up the findings and to answer the research question of the study, it can be said that content analysis of the raters' verbal protocols showed that neither increase nor decrease in test length made any noticeable differences in the raters' mental processes while scoring writing tests.

Discussion

To summarize the results of the present study, it can be said that the length of the writing test did not result in any significant differences in the mental processes through which raters assigned scores to the test-takers' writing performance. In other words, the raters scored the three versions of the writing test (the original, longer, and shorter versions) with a similar viewpoint towards the important elements of writing.

The results of this study are in line with other studies in the literature. In a similar study, Hoora (2019) investigated the effect of the duration of the IELTS speaking test on the examiners' evaluation of the candidates' performance. Similar to the results of this study, it was found that, although there were certain factors in the band descriptors of the IELTS speaking test the raters paid more attention to while ignoring some other factors in rating three versions of the IELTS speaking test (i.e., the original, longer, and shorter versions), their mental processes in scoring the three versions were not significantly different from each other. In addition, there was not a considerable difference in the scores they assigned to the three versions of the speaking test. The similar results of this study and Hoora's show that test length, whether in writing or speaking tests, does not make a significant difference in the raters' mental processes when scoring the test-takers' performance.

Examining the raters' preferences in scoring the test-takers' performance, Van Batenburg et al. (2018) researched the examiners' judgment of the speaking performance of learners in different task types where the raters scored the participants' performance both holistically and analytically using two different scoring rubrics. They found that there was a high correlation between the scores

given by the three raters on both holistic and analytic scales in measuring different task types. In other words, the scores assigned by the three raters to different task types were not significantly different from each other following either scale. These results are somehow in line with the results of the present study, showing the consistency in the raters' scoring in different situations.

Considering the differences in the raters' scoring of writing tests, Attali (2016) conducted a study in which the performance of newly-trained and experienced raters was compared. He came up with the conclusion that the scores provided by the newly-trained raters were very much similar to the ones by the experienced raters. In other words, the training turned out to be an influential factor in increasing the correlations between the raters' scores. In fact, training resulted in a beneficial decrease in the difference between the newly-trained and experienced raters' scoring of writing. Thus, training, in contrast to test length, made a difference in the rater's scoring of writing performance.

Overall, the comparison of the results of the present study with other studies showed that raters, their mental processes, and their characteristics (e.g., experience) can be considered as important factors when scoring speaking and/or writing performance. Regarding the raters' mental processes, which was the focus of the present study, the interesting finding was that the length of the writing test was not a determining factor for the raters and they paid more attention to the criteria in the writing scoring rubric regardless of the test length. The differences in the tasks or tests may not make much differences in the scores assigned by the raters or the mental processes they engage in as the results of this study showed. In other words, variations in task types, the word count (or test length), the time allotted to do the task, and related issues might not have a noteworthy effect on the scores assigned by the raters. However, the raters' individual characteristics, such as experience, training, and related factors might be probable influential factors.

Conclusion

The goal of this study was to investigate the effect of test length on the raters' mental processes while scoring candidates' writing performance where both groups of participants (i.e., test-takers and raters) were chosen based on availability sampling. In general, it was found that raters considered certain elements of writing

more important and disregarded some other elements when assigning scores to a piece of writing regardless of its length.

The main limitation of the study was the few number of test-takers and raters, which reduced the generalizability of the findings. Therefore, the results of the study should be interpreted cautiously. Further, the major delimitation of the study was that both test-takers and raters were female since we wanted to control the gender effect in this study. However, using a combination of male and female test-takers and raters makes comparisons across the different groups possible and may result in more valuable findings in the future.

The results of this study can be useful for raters since, to be a fairer rater who assigns more reliable scores, it is necessary to be aware of one's own characteristics, to go to routine training sessions, to gain more experience, to get familiar with different types of scoring rubrics, and so forth. The findings of the present study should make raters aware of different important points to take into consideration when scoring a piece of writing following a specific scoring rubric in addition to the common and typical points.

Rater trainers can also benefit from the results of the present study since it makes them aware of some important points raters do not usually pay attention to in scoring writing. Although *answering the question*, *format of the essay*, *avoiding irrelevant details*, *aspects of cohesion*, and *referencing* are some of the essential points that need more attention on the part of the raters, it was found in the present research that raters often ignore them. Therefore, rater trainers should emphasize such points more when training the raters.

Finally, interested researchers are invited to conduct follow-up research where the personality type of raters, their gender, their experience, and their training are considered to see if such differences result in the same or different mental processes raters engage in when assigning scores to different types of writing and/or speaking tasks or tests.

References

Ackerman, P. L., & Kanfer, R. (2009). Test length and cognitive fatigue: An empirical examination of effects on performance and test-taker reactions. *Journal of Experimental Psychology: Applied, 15*(2), 163-181. https://doi.org/10.1037/a0015719

Ahmad, B. A. D. (2021). Effect of time allotment on test scores for academic writing of Indonesian learners of English. *Multicultural Education, 7*(1), 134-141.

Alderson, J. C., & Banerjee, J. (2002). Language testing and assessment (Part 2). *Language Teaching, 35*(2), 79-113. https://doi.org/10.1017/S0261444802001751

Asadi Vahdat, Z., & Tavassoli, K. (2019). A comparison of the effects of task repetition and elicitation techniques on EFL learners' expository and descriptive writing. *Language Horizons, 3*(1), 243-267. https://doi.org/10.22051/lghor.2019.27890.1173

Attali, Y. (2016). A comparison of newly-trained and experienced raters on a standardized writing assessment. *Language Testing, 33*(1), 1-17. https://doi.org/10.1177/0265532215582283

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford University Press.

Barkaoui, K. (2019). Examining sources of variability in repeaters' L2 writing scores: The case of the PTE Academic writing section. *Language Testing, 36*(1), 3-25. https://doi.org/10.1177/0265532217750692

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45*(1), 5-35. https://doi.org/10.5054/tq.2011.244483

Bijani, H. (2018). The investigation of rater expertise in oral language proficiency assessment: A multifaceted Rasch analysis. *Language Horizons, 2*(2), 103-124. https://doi.org/10.22051/lghor.2019.26072.1123

Brown, J. D. (2005). *Testing in language programs* (2nd ed.). McGraw-Hill College.

Coe, K., & Scacco, J. M. (2017). Content analysis, quantitative. In C. S. Davis & R. F. Potter (Eds.), *The international encyclopedia of communication research methods* (pp. 1-11). John Wiley & Sons, Inc.

Cohen, A. D. (2003). Learner strategy training in the development of pragmatic ability. In A. Martinez Flor, E. Usó Juan, & A. Fernández Guerra (Eds.), *Pragmatic competence and foreign language teaching* (pp. 93-108). Publicacions de la Universitat Jaume I.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies.* Oxford University Press.

Ducasse, A. M. (2010). *Interaction in paired oral proficiency assessment in Spanish: Rater and candidate input into evidence-based scale development and construct definition.*

Peter Lang.

Duijm, K., Schoonen, R., & Hulstijn, J. H. (2018). Professional and non-professional raters' responsiveness to fluency and accuracy in L2 speech: An experimental approach. *Language Testing, 35*(4), 501-527. https://doi.org/10.1177/0265532217712553

Eckes, T. (2012). Operational rater types in writing assessment: Linking rater cognition to rater behavior. *Language Assessment Quarterly, 9*(3), 270-292. https://doi.org/10.1080/15434303.2011.649381

Esfandiari, R., & Noor, P. (2019). Iranian EFL raters' cognitive processes in rating IELTS speaking tasks: The effect of expertise. *Journal of Modern Research in English Language Studies, 5*(2), 41-76. https://doi.org/10.30479/jmrels.2019.9383.1248

Fazilatfar, A., Kasiri, F., & Nowbakht, M. (2020). The comparative effects of planning time and task conditions on the complexity, accuracy, and fluency of L2 writing by EFL learners. *Iranian Journal of Language Teaching Research, 8*(1), 93-110.

Han, Q. (2016). Rater cognition in L2 speaking assessment: A review of the literature. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics, 16*(1), 1-24. https://doi.org/10.7916/D82R53MF

Hoora, E. (2019). *The effect of the duration of IELTS speaking test on examiners' evaluation of candidates' performance* [Unpublished Master's thesis]. Karaj Islamic Azad University, Iran.

Huang, J., & Foote, C. J. (2010). Grading between the lines: What really impacts professors' holistic evaluation of ESL graduate student writing? *Language Assessment Quarterly, 7*(3), 219-233. https://doi.org/10.1080/15434300903540894

Humphrey-Murto, S., Shaw, T., Touchie, C., Pugh, D., Cowley, L., & Wood, T.J. (2021). Are raters influenced by prior information about a learner? A review of assimilation and contrast effects in assessment. *Advances in Health Sciences Education, 25*(2), 4-24. https://doi.org/10.1007/s10459-021-10032-3

Kang, H. S., & Veitch, H. (2017). Mainstream teacher candidates' perspectives on ESL writing: The effects of writer identity and rater background. *TESOL Quarterly, 51*(2), 249-274. https://doi.org/10.1002/tesq.289

Llosa, L., & Malone, M. E. (2019). Comparability of students' writing performance on TOEFL iBT and in required university writing courses. *Language Testing, 36*(2), 235 263.

Lumley, T. (2005). *Assessing second language writing: The rater's perspective.* Peter Lang.

Marcoulides, G. A., & Ing, M. (2014). The use of generalizability theory in language assessment. In A. J. Kunnan (Ed.), *The companion to language assessment: Evaluation, methodology and interdisciplinary themes* (Vol. 3, pp. 124-141). John Wiley & Sons, Inc.

May, L. (2011). Interactional competence in a paired speaking test: Features salient to raters. *Language Assessment Quarterly, 8*(2), 127-145. https://doi.org/10.1080/15434303.2011.565845

Oh, S. (2020). Second language learners' use of writing resources in writing assessment. *Language Assessment Quarterly, 17*(1), 60-84. https://doi.org/10.1080/15434303.2019.1674854

Plakans, L. (2014). Written discourse. In A. J. Kunnan (Ed.), *The companion to language assessment: Evaluation, methodology and interdisciplinary themes* (Vol. 3, pp. 305317). John Wiley & Sons, Inc.

Purpura, J. E. (2004). *Assessing grammar.* Cambridge University Press.

Purpura, J. E. (2014). Cognition and language assessment. In A. J. Kunnan (Ed.), *The companion to language assessment* (Vol. 3, pp.1452-1476). John Wiley & Sons, Inc.

Sahin, A., & Anil, D. (2017). The effects of test length and sample size on item parameters in 1452item response theory. *Educational Sciences: Theory & Practice, 17*(1), 321-335. https://doi.org/10.12738/estp.2017.1.0270

Sahragard, R., & Mallahi, O. (2014). Relationship between Iranian EFL learners' language learning styles, writing proficiency, and self-assessment. *Social and Behavioral Sciences, 98*(1), 1611-1620. https://doi.org/10.1016/j.sbspro.2014.03.585

Sawaki, Y. (2007). Construct validation of analytic rating scales in a speaking assessment: Reporting a score profile and a composite. *Language Testing, 24*(3), 355-390. https://doi.org/10.1177/0265532207077205

Shirai, Y., & Vercellotti, M. L. (2014). Language acquisition and language assessment. In A. J. Kunnan (Ed.), *The companion to language assessment: Evaluation, methodology and interdisciplinary themes* (Vol. 3, pp. 300-314). John Wiley & Sons, Inc.

University of Cambridge ESOL examinations. (2015). *Cambridge English IELTS 10 with answers: Authentic examination papers from Cambridge English language assessment*. Cambridge University Press.

Van Batenburg, E. S., Oostdam, R. J., Van Gelderen, A. J., & De Jong, N. H. (2018). Measuring L2 speakers' interactional ability using interactive speech tasks. *Language Testing, 35*(1), 75-100. https://doi.org/10.1177/0265532216679452

Weigle, S. C. (2002). *Assessing writing.* Cambridge University Press.

Wind, S. A. (2019). A nonparametric procedure for exploring differences in rating quality across test-taker subgroups in rater-mediated writing assessments. *Language Testing, 36*(4), 595-616. https://doi.org/10.1177/0265532219838014

Paper Abstracts in Persian

مجله افق‌های زبان
LGHOR
دانشگاه الزهرا

فصلنامهٔ علمی افق‌های زبان دانشگاه الزهرا(س)
سال هفتم، شمارهٔ ۱، بهار ۱۴۰۱

بررسی رابطه بین مقیاس‌های بار شناختی ذهنی و عینی دشواری سؤال‌های زبان انگلیسی

مقاله پژوهشی
صفحات ۳۳-۷

شادی نوروزی[۱]
حسین کرمی[۲*]
زری سعیدی[۳]

تاریخ دریافت: ۱۴۰۰/۰۷/۰۹ تاریخ تصویب: ۱۴۰۰/۱۲/۲۵

چکیده

پژوهش حاضر به بررسی رفتار پاسخ گویی و ادراک آزمون‌دهندگان، هنگام پاسخ‌گویی به سؤال‌های یک آزمون، از دیدگاه نظریه بار شناختی پرداخت. داده‌های تجربی از ۶۰ دانش‌آموخته و دانشجوی کارشناسی ارشد زبان انگلیسی با سطح بالایی از مهارت زبان گردآوری شدند. شرکت‌کنندگان می‌بایست به ۶۰ سؤال زبان چندگزینه‌ای (یعنی سؤال‌های دستور زبان و واژگان)، که از آزمون‌های سرنوشت‌ساز ورودی دانشگاه ایران برای گرایش‌های کارشناسی ارشد زبان انگلیسی گرفته شده بودند، به سریع‌ترین و دقیق‌ترین شکل ممکن پاسخ می‌دادند. پس از تکمیل هر سؤال، آن‌ها درک خود را از نظر دشواری سؤال‌های آزمون (برتفیش و همکاران، ۱۹۷۲) و میزان تلاش ذهنی (پاس، ۱۹۹۲) ارزیابی کردند. زمان پاسخ‌گویی آن‌ها به هر سؤال زبان و گزینه‌های انتخابی آن‌ها نیز توسط نرم افزار سایکوپای ذخیره شد (پیرس و همکاران، ۲۰۱۹). یافته‌ها، با استفاده از همبستگی‌های پیرسون و اسپیرمن رو، نشان دادند که زمان پاسخ‌گویی با میزان تلاش ذهنی همبستگی مثبت قوی دارد، بدین معنا که هر دو مقیاس عینی و ذهنی بار شناختی از نظر حساسیت به تغییرات بار شناختی در سؤال‌های آزمون زبان منطبق هستند. افزون بر این، مقیاس‌های تلاش ذهنی ادراک‌شده و سطح دشواری ادراک‌شده نشان دادند که این دو شاخص‌های درست ارزیابی تغییرات بار شناختی هستند. همان گونه که پیش‌بینی شده بود، زمان پاسخ‌گویی نیز نشان داد که سؤال‌های دشوارتر آزمون زبان، بار بیشتری را بر آزمون دهندگان تحمیل می کنند.

کلیدواژه‌ها: بار شناختی، سؤال‌های زبان انگلیسی، سؤال‌های چهار گزینه‌ای، زمان پاسخ‌گویی، مقیاس‌های عینی/ذهنی

---

* نویسندهٔ مسئول

[۱] دانش‌آموختهٔ کارشناسی ارشد، گروه زبان انگلیسی، دانشکده زبان و ادبیات خارجی، دانشگاه تهران، تهران، ایران. shadi.noroozi@ut.ac.ir

[۲] استادیار، گروه زبان انگلیسی، دانشکده زبان‌ها و ادبیات خارجی، دانشگاه تهران، تهران، ایران. hkarami@ut.ac.ir

[۳] دانشیار، گروه زبان و ادبیات انگلیسی، دانشکده ادبیات فارسی و زبان‌های خارجی، دانشگاه علامه طباطبایی، تهران، ایران. saeedi.za@atu.ac.ir

فصلنامهٔ علمی افق‌های زبان دانشگاه الزهرا(س)
سال هفتم، شمارهٔ ۱، بهار ۱۴۰۱

تأثیر آموزش خود-تنظیمی بر درک مطلب و تمایل به پذیرش اشتباه در کلاس‌های حضوری در مقایسه با کلاس‌های برخط

مقاله پژوهشی
صفحات ۵۸-۳۵

محمدرضا زبردست[۱]
مهری حداد نارافشان[۲*]
حسن شهابی[۳]

تاریخ ارسال: ۱۴۰۰/۰۴/۰۸ تاریخ پذیرش: ۱۴۰۰/۱۱/۲۳

چکیده

فناوری‌های نوظهور و کاربردهای آن‌ها در آموزش، ما را ملزم به تلاش در ایجاد و آزمونِ نظریه‌هایی می کنند که می‌توانند بر بهره‌مندی بهینه یادگیری دانش‌آموز و آموزش معلم تأثیر بگذارند. نظریه یادگیری خود-تنظیمی، یکی از تأثیرگذارترین نظریه‌هایی است که پیوندهای عوامل مختلف برای یادگیری بهینه به مثابه یک فرایند در حال پیشرفت را شناسایی می‌کند. بنابراین، این پژوهش سعی کرده است کارایی یادگیری خود-تنظیمی در رشد درک مطلب زبان‌آموز و تمایل به پذیرش اشتباه در محیط یادگیری برخط را افزون بر محیط یادگیری حضوری بررسی کند. برای پرداختن به این موضوع، از یک طرح شبه-آزمایشی استفاده شد. به گروه‌های آزمایشی تعلیم داده شد که از طریق دریافت راهبردهای یادگیری خودتنظیمی به قابلیت خود-تنظیمی برسند در حالی که گروه کنترل هیچ‌گونه دستورالعمل خود-تنظیمی دریافت نکردند. تجزیه و تحلیل آماری آزمون تی-جفتی نشان داد که فقط گروه‌های آزمایشی پیشرفت قابل توجهی در درک مطلب خواندن و تمایل به پذیرش اشتباه داشتند. افزون بر این، همان گونه که تجزیه و تحلیل آماری آزمون کوواریانس نشان داد، گروه برخط در مقایسه با گروه کنترل و حضوری در زمینه مهارت درک مطلب خواندن عملکرد بهتری داشت. این یافته‌ها می‌تواند مورد توجه پژوهشگران، معلمان و طراحان دوره‌های آموزشی به منظور اهمیت استفاده از مدل‌های خود-تنظیمی به عنوان روشی جایگزین برای آموزش سنتی قرار بگیرد.

کلیدواژه‌ها: کلاس حضوری، کلاس برخط، درک مطلب خواندن، خود-تنظیمی، تمایل به پذیرش اشتباه

---

* نویسندهٔ مسئول
[۱] دانشجوی دکترا، گروه زبان های خارجی، دانشکده ادبیات و علوم انسانی، واحد کرمان، دانشگاه آزاد اسلامی، کرمان، ایران.
aseman2st@gmail.com
[۲] استادیار، گروه زبان‌های خارجی، دانشکده ادبیات و علوم انسانی، واحد کرمان، دانشگاه آزاد اسلامی، کرمان، ایران.
haddad@iauk.ac.ir
[۳] استادیار، گروه زبان های خارجی، دانشکده ادبیات و علوم انسانی، واحد کرمان، دانشگاه آزاد اسلامی، کرمان، ایران.
hshahabi@iauk.ac.ir

مجله افق‌های زبان
LGHOR
دانشگاه الزهرا

فصلنامهٔ علمی افق‌های زبان دانشگاه الزهرا(س)
سال هفتم، شمارهٔ ۱، بهار ۱۴۰۱

عملکرد نویسندگان متون انگلیسی به عنوان زبان دوم در موقعیت‌های نگارش دانشگاهی در شرایط آزمون و در شرایط غیر آزمون

مقاله پژوهشی
صفحات ۸۵-۵۹

فهیمه معرفت[*۱]
مجتبی حیدری[۲]

تاریخ دریافت: ۱۴۰۰/۰۶/۰۷ تاریخ پذیرش: ۱۴۰۰/۱۱/۲۳

چکیده

با گسترش دسترسی به انواع ابزارهای ارجاعی جدید، نویسندگان متون انگلیسی به عنوان زبان دوم امروزه انتخاب‌های متنوعی برای نگارش تکالیف دانشگاهی خود دارند. نبود چنین مرجع‌هایی در بسیاری از آزمون‌های سرنوشت‌ساز دانشگاهی نوشتاری باعث ایجاد تفاوت در دو موقعیت می‌شود. پژوهش حاضر با هدف مقایسه تکلیف‌های نویسندگان متون انگلیسی به عنوان زبان دوم در دو موقعیت الف) نوشتاری در شرایط امتحان و ب) در شرایط غیر امتحانِ زندگی واقعی، هفت دانشجوی زبان انگلیسی به عنوان زبان دوم که قبلاً آزمون آیلتس داده بودند را انتخاب کرد. تحلیل ویژگی‌های زبانیِ مشخصه‌های زبان‌شناختیِ پیچیدگی نحوی، پیچیدگی واژگانی و انسجام متن نوشتاری نشان می‌دهد که متن تولیدی دانشجویان در موقعیت نگارش دانشگاهی زندگی واقعی از کیفیت بالاتری برخوردار است. دانش‌آموزان دو تکلیف نوشتاری مشابه را انجام دادند: یک آزمون موقعیت-امتحان و یک آزمون نوشتار غیر امتحان که هدف آن شبیه‌سازی محیط زندگی واقعی بود. افزون بر این، تحلیل اف.ای.اس.تی.اس کیفیت نوشتارها که توسط ارزیابی‌کنندگان انسانی ارزیابی شد، نشان داد که دانشجویان به صورت برابر از مزایای محیط‌های زندگی واقعی در مقایسه با موقعیت‌های امتحان بهره‌مند نیستند. یافته‌ها همچنین نشان می‌دهد که دانشجویان هنگام نگارش، مدت زمان متفاوتی را صرف نموده و از انواع مختلف جست‌وجو برای کسب راهنمایی از منابع خارجی استفاده کردند. پیشینهٔ آموزشی دانشجویان و راهبردهای نگارشی آن‌ها می‌تواند عملکرد آن‌ها در زندگی واقعی نگارش دانشگاهی در مقایسه با نگارش آزمون را تحت تاثیر خود قرار دهد، نکته‌ای که هشداری است در زمینه روایی آزمون‌های سرنوشت‌ساز دانشگاهی.

کلیدواژه‌ها: نوشتار آکادمیک، آزمون ویژگی‌های زبانی، موقعیت‌های غیرامتحان، اف.ای.اس.تی.اس ، روایی امتحان نوشتاری

---

* نویسندهٔ مسئول

[۱] استاد، گروه زبان و ادبیات انگلیسی، دانشکده ادبیات فارسی و زبان‌های خارجی، دانشگاه علامه طباطبایی، تهران، ایران. fmarefat@atu.ac.ir

[۲] دانشجوی دکترا، گروه زبان و ادبیات انگلیسی، دانشکده ادبیات فارسی و زبان‌های خارجی، دانشگاه علامه طباطبائی، تهران، ایران. Heydari_mojtaba@atu.ac.ir

فصلنامهٔ علمی افق‌های زبان دانشگاه الزهرا(س)
سال هفتم، شمارهٔ ۱، بهار ۱۴۰۱

ایدئولوژی و امر متعالی در مقدمه ویلیام وردزورث

مقاله پژوهشی
صفحات ۱۰۶-۸۷

غیاث‌الدین علیزاده[*۱]
مسعود فرهمندفر[۲]

تاریخ ارسال: ۱۴۰۰/۰۴/۲۹ تاریخ پذیرش: ۱۴۰۰/۱۱/۲۳

چکیده

برای درک حدود و ثغور رشدِ ایدئولوژیکی وُردزورت و نقش امر متعالی در توسعهٔ ذهن شاعر، توصیه می‌شود پیشرفت زمانی و روایی *مقدمه* (۱۷۹۹)، را دنبال کنیم. مقالهٔ حاضر، از طریق تحلیل آرای ژیژکی *شاهکار بزرگ* وردورت، نشان می‌دهد که می‌توان به امر متعالی از دو منظر مختلف دست یافت. نخست، می‌توان آن را مداخلهٔ نمادین در درون جهان خیالیِ سوژه دانست که حس وحدت و کمال را دچار تشویش می‌کند و اُبژهٔ تمنا را از ترتیب خیالی به نمادین تغییر می‌دهد و دوم اینکه، می‌توان آن را مداخلهٔ امر واقع در ترتیب نمادین دانست، به منزلهٔ فقدان در دل دیگری، که در نتیجه آن، هم‌خوانیِ کاملِ عناصر دلالتی و دست‌یابی نهایی به اُبژهٔ تمنّا غیر ممکن می‌شود. آن‌چه این دو دلالت متفاوت از امر متعالی را به هم پیوند می‌زند، درک این مهم است که در هر دو مورد، امر متعالی جایگاه حالت‌های وهم‌گون و ناموجود را دارد.

کلیدواژه‌ها: وردزورث، *مقدمه*، امر متعالی، ژیژک، ادبیات رمانتیک

* نویسنده مسئول
[۱] استادیار، گروه زبان و ادبیات انگلیسی، دانشکده ادبیات و علوم انسانی، دانشگاه ملایر، ملایر، ایران.
ghiasuddin.alizadeh@malayeru.ac.ir
[۲] استادیار، گروه زبان و ادبیات انگلیسی، دانشکده ادبیات فارسی و زبان‌های خارجی، دانشگاه علامه طباطبائی، تهران، ایران.
farahmand@atu.ac.ir

مجله افق‌های زبان
LGHOR
دانشگاه الزهرا
فصلنامهٔ علمی افق‌های زبان دانشگاه الزهرا(س)
سال هفتم، شمارهٔ ۱، بهار ۱۴۰۱

آموزش صورت-محور فعال در کلاس آموزش معکوس: رشد و ماندگاری دانش ضمنی و آشکار دستور زبان

مقاله پژوهشی
صفحات ۱۳۴-۱۰۷

احمد خلیفه[۱]
محمد بوالی[۲*]
احسان رسایی[۳]

تاریخ دریافت: ۱۴۰۰/۰۷/۱۲ تاریخ پذیرش: ۱۴۰۰/۰۹/۲۸

چکیده

مطالعهٔ حاضر به بررسی تأثیرات آموزش صورت-محور فعال بر بهبود دانش صریح و ضمنی دستور زبان فراگیران زبان انگلیسی ایرانی در یک کلاس سنتی و یک کلاس آموزش معکوس می‌پردازد، همچنین تأثیر کلاس آموزش معکوس بر ماندگاری دانش آشکار و ضمنی بررسی می‌شود. دو کلاس به صورت تصادفی به عنوان گروه آزمایشی (۳۱ نفر) و گروه کنترل (۲۸ نفر) انتخاب شدند. گروه نخست آموزش متمرکز بر صورت فعال از طریق شرح فرازبانی در کلاس آموزش معکوس را دریافت نمودند در حالی که گروه دوم در یک کلاس سنتی حضور یافتند. برای ارزیابی پیشرفت و ماندگاری دانش صریح و ضمنی دستور زبان فراگیران، از دو آزمون دانش ضمنی و دانش آشکار استفاده شد. یک مجموعه پیش‌آزمون و دو مجموعه پس‌آزمون بلافاصله و چهار هفته پس از جلسهٔ آموزش اجرا شد. دو تحلیل واریانس چندمتغیره ۳×۲ مختلط و آزمون‌های *تعقیبی* برای کشف تفاوت‌های میان عملکرد بین گروهی و درون گروهی به اجرا درآمد. یافته‌های پژوهش نشان داد که هر دو کلاس به یک اندازه در دانش آشکار و ضمنی دستور زبان فراگیران بهبود یافتند. همچنین، گروه آزمایشی میزان بیشتری از ماندگاری دانش را نشان داد. این یافته‌ها می‌تواند به طراحان برنامه‌های آموزش زبان، سیاست‌گذاران عرصه آموزش و معلمان زبان کمک نماید و اطمینان دهد تا با بهره‌گیری از دوره‌های آموزش معکوس، آموزش دستور زبان انگلیسی را به شیوه مؤثری انجام داده و به زبان‌آموزان کمک نمایند دانش آشکار و ضمنی دستور زبان را به مدت طولانی‌تری حفظ نمایند.

کلیدواژه‌ها: دانش آشکار، آموزش معکوس، صورت-محور، دانش ضمنی، ماندگاری

---

* نویسندهٔ مسئول
[۱] دانشجوی دکتری، گروه زبان انگلیسی، واحد شیراز، دانشگاه آزاد اسلامی، شیراز، ایران.
ahmadkhalifeh@iaubushehr.ac.ir
[۲] استادیار ،گروه زبان انگلیسی، واحد شیراز، دانشگاه آزاد اسلامی، شیراز، ایران. mbvl57@gmail.com
[۳] دانشیار، گروه زبان انگلیسی، واحد شیراز، دانشگاه آزاد اسلامی، شیراز، ایران. ehsan.rassaei@majancollege.edu.om

فصلنامهٔ علمی افق‌های زبان دانشگاه الزهرا(س)
سال هفتم، شمارهٔ ۱، بهار ۱۴۰۱

تأثیر رشد راهبرد خود-تنظیمی بر توانایی فراگیران زبان انگلیسی سطح متوسط در خواندن خودتنظیمی

مقاله پژوهشی
صفحات ۱۵۷-۱۳۵

محبوبه مرشدیان[1*]

تاریخ دریافت: ۱۴۰۰/۰۵/۰۹ تاریخ پذیرش: ۱۴۰۱/۰۱/۰۳

چکیده

خود-تنظیمی آکادمیک فرایندی است که در آن زبان‌آموزان از بازخورد شناختی، عاطفی، رفتاری و انگیزشی برای تنظیم یا اصلاح رفتارها و راهبردها برای دست‌یابی به هدف‌های خود استفاده می‌کنند. به این ترتیب، یادگیری خود-تنظیم‌یافته فرایندی در تعیین اهداف، برنامه‌ریزی راهبردی، انتخاب و استفاده از راهبردها، نظارت بر کارامدی فرد و ارزیابی خویش است. بر این مبنا، این پژوهش به تأثیر آموزش خود-تنظیمی بر اساس راهبرد تفکر قبل، حین و بعد خواندن از طریق توسعه راهبرد خود-تنظیم یافته بر خود-تنظیمی فراگیران انگلیسی سطح متوسط در خواندن انگلیسی به عنوان زبان دوم متن‌های توضیحی می‌پردازد. توسعه راهبرد خود-تنظیم‌یافته، برای راهبرد تفکر قبل و حین و بعد خواندن در جلسه‌های خواندن زبان انگلیسی به عنوان زبان دوم گروه آزمایشی دست‌نخورده پیاده‌سازی شد، ولی به گروه کنترل دست‌نخورده آموزش معمول خواندن زبان انگلیسی به عنوان زبان دوم داده شد، با این هدف که آنها متون را بخوانند و به سؤال‌های درک مطلب آن پاسخ دهند. داده‌های خود-تنظیمی با استفاده از پرسشنامه پارامتری راهبردهای انگیزشی تنظیم‌شده برای یادگیری قبل و بعد از آموزش گردآوری شد. تحلیل کوواریانس یک‌طرفه نشان‌داد که آموزش خود-تنظیمی مبتنی بر راهبرد تفکر قبل و حین و بعد خواندن از طریق رشد راهبرد خود-تنظیم‌یافته می‌تواند مهارت خواندن خود-تنظیمی زبان‌آموزان را تقویت‌کند. این یافته‌ها می‌تواند معلمان را ترغیب کند که در آموزش زبان انگلیسی به زبان آموزان در رشد راهبرد خود-تنظیمی برای راهبرد تفکر قبل و حین و بعد خواندن و کسب توانایی خود-تنظیمی فرایند خواندن زبان انگلیسی کمک کنند.

کلیدواژه‌ها: خود-تنظیمی، توسعه راهبرد، خواندن، یادگیری خود-تنظیم‌یافته، فراگیران سطح متوسط زبان انگلیسی

[1] استادیار، گروه مترجمی زبان، موسسه آموزش عالی طلوع مهر، قم، ایران؛ m.morshedian@tolouemehr.ac.ir

مجله افق‌های زبان
LGHOR
دانشگاه الزهرا

فصلنامهٔ علمی افق‌های زبان دانشگاه الزهرا(س)
سال هفتم، شمارهٔ ۱، بهار ۱۴۰۱

تأثیر طول آزمون بر فرایندهای ذهنی ارزیاب‌ها هنگام نمره‌دهی به عملکرد نوشتاری آزمون‌دهندگان

مقاله پژوهشی
صفحات ۱۸۲-۱۵۹

فاطمه نیک مراد[۱]
کبری توسلی[۲*]

تاریخ ارسال: ۱۴۰۰/۰۵/۲۸ تاریخ پذیرش: ۱۴۰۰/۱۱/۲۳

چکیده

عوامل مختلفی از جمله ژانر نگارش، داستان‌پردازی و یا طول آزمون می‌توانند بر فرایندهای ذهنی مصححین هنگام نمره‌دهی به آزمون‌های نگارش تأثیر بگذارند. در این راستا، اینکه آیا افزایش یا کاهش طول آزمون تأثیری بر چگونگی سنجش ارزیاب‌ها از عملکرد نوشتاری آزمون‌دهندگان دارد، انگیزه اصلی انجام این پژوهش بود. بدین منظور، ۱۲ فراگیر زبان انگلیسی که در بخش نگارش آزمون آزمایشی آیلتس نمره بین ۵/۵ تا ۷/۵ دریافت کرده بودند، از طریق نمونه‌گیری در دسترس انتخاب شدند. شرکت‌کننده‌ها سه مقاله استدلالی به صورت نسخه اصلی، بلندتر و کوتاه‌تر نوشتند. هر یک از سه نسخه آزمون‌دهنده گان، سپس توسط سه ارزیاب با استفاده از مقیاس تصحیح نگارش تکلیف دوم آیلتس تصحیح شد. در این میان، ارزیاب‌ها پروتکل‌های شفاهی هم ارائه دادند که در آن‌ها به توضیح جزئی دلایل اصلی‌شان برای اختصاص نمره‌ها به هر مقاله پرداختند. سپس، پروتکل‌های شفاهی به نوشته بازگردانده شدند و محتوای آن‌ها با استفاده از نرم‌افزار ان.وی.وو نسخه ۱۱ تجزیه و تحلیل شد تا مضمون های اصلی ذکرشده توسط ارزیاب‌ها در نمره‌دهی به هر آزمون نوشتاری استخراج شود. یافته‌ها نشان داد که ارزیاب‌ها توجه ویژه‌ای به برخی از معیارهای موجود در مقیاس تصحیح نگارش آیلتس داشتند و برخی عوامل را نادیده گرفتند. با این وجود، یک الگوی مشابه در میان همه ارزیاب‌ها در نمره‌دهی سه آزمون نوشتاری وجود داشت. یافته‌ها نشان‌دهنده وجود تفاوت‌های معنادار در فرایندهای ذهنی ارزیاب‌ها هنگام نمره‌دهی به هر سه آزمون نگارش نبود. نتیجه این پژوهش نشان داد طول آزمون عامل تأثیرگذاری بر فرایندهای ذهنی ارزیاب‌ها در آزمون‌های نگارشی نبود. در نتیجه، ارزیاب‌ها و طراحان آزمون نیازی نیست نگران تأثیرگذاری طول آزمون بر نمره‌دهی ارزیاب‌ها باشند.

کلیدواژه‌ها: فرایندهای ذهنی، ارزیاب، طول آزمون، پروتکل شفاهی، آزمون نگارش

---

* نویسندهٔ مسئول

[۱] دانشجوی دکترا، گروه زبان انگلیسی، دانشکده ی ادبیات و زبانهای خارجی، واحد کرج، دانشگاه آزاد اسلامی، کرج، ایران.
fateme.nikmard@kiau.ac.ir

[۲] استادیار، گروه زبان انگلیسی، دانشکده ی ادبیات و زبانهای خارجی ، واحد کرج، دانشگاه آزاد اسلامی، کرج، ایران.
kobra.tavassoli@kiau.ac.ir

فصلنامهٔ علمی افق‌های زبان دانشگاه الزهرا(س)
سال هفتم، شمارهٔ ۱، بهار ۱۴۰۲

فهرست مطالب

معاونت پژوهشی دانشگاه

فصلنامهٔ علمی

افق‌های زبان دانشگاه الزهرا(س)

سال هفتم، شماره یکم، بهار ۱۴۰۲ (پیاپی ۱۵)

صاحب امتیاز: **دانشگاه الزهرا(س)**

مدیر مسئول: **دکتر محسن شیرازی‌زاده**

سردبیر: **دکتر محمدرضا عنانی سراب**

ویراستار زبان انگلیسی: **دکتر شهره تیمورنژاد**

ویراستاران زبان فارسی: **دکتر نرجس منفرد**

مدیر اجرایی: **نرگس جعفری**

اعضای هیئت تحریریه

آدرس: تهران، ونک، دانشگاه الزهرا(س)، دانشکدهٔ ادبیات، کد پستی: ۱۹۹۳۸۹۱۱۷۶

پست الکترونیکی: lghorizons@alzahra.acir

شاپا: ۳۵۰x-۲۵۸۸

شاپای الکترونیکی: ۵۶۳۴-۲۵۸۸

مجلهٔ افق های زبان با همکاری انجمن زبان شناسی ایران منتشر می‌شود.

LGHOR

انجمن زبان‌شناسی ایران