A Comparison of Paper-Based Tests and Computer-Based Tests in the Emergence of COVID- 19 Era: Evidence by Event-Related Potentials

Barani, reyhaneh*
PhD CandidateAzadshahrblv, MASHHAD, IRA.

ABSTRACT

*R*eading comprehension is one of the most important sections of the international tests including IELTS or TOFEL. Also, reading comprehension tests employed at schools or university are attempting to determine the comprehension level of the test takers specifically inCorona virus emergence era. Metacognitive strategies are used by test takers in order to gain their intended marks in tests rather than comprehending the text. Thus, it is a tool for readers to increase their decoding skills. This study aimed at exploring the comprehension level of the test takers without the effects of Metacognitive strategies. While there seems to be a deep linkage between the brain and reading comprehension, Event-related potentials (ERPs) as one of the methods of brain activity measuring allows researchers to observe reading-related brain processes and can document neural patterns at the millisecond level. This study aimed at high-lightening the role of computer-based testing through ERP. 10 Iranian IELTS candidates holding the band scores from 6 to 8 participated in this study. The reading comprehension items of a retired version of IELTS were given to the test takers as the paper-version test and as the stimuli in computer-version test format. In this research, mixed method was used. This study aimed at comparing the paper-version and computer-based tests. EEG signals of the participants were recorded during the computer-based version test and ERPs were extracted. Findings showed that computer-version testing can examine the readers' comprehension level directly.*

Keywords: Event-related potentials (ERPs), reading comprehension, computer-version testing, paper-version testing, Metacognitive strategies

Introduction

Among all the four skills in language, scholars pay particular attention to reading comprehension skill and they believe that it is a criterion for determining whether a person is academically literate or not (Kovalenko, 2018; Moore et al., 2012). Reading comprehension skill in comparison to the other language skills is considered as a key element for educationists, psychologists, and test designers (Lui,2010). Test designers of reading comprehension section in different widespread language tests such as IELTS and TOEFLE have measured this skill by using paper based tests, questionnaires, self-reports, and subjective ratings. Thus, they should give more prominence to designing a comprehensive computer based test in order to make the results closer to the aims of the framework that is defined for each exam (Alderson, 2000; Urquhart & Weir, 2014, cited in Kovalenko, 2018).Computer based tests by using the technology of Event-related potentials can be a powerful tool to directly measure the hidden processes underlying reading comprehension skill(Rastelli,2018).The brain is linked with every language skill, butthe researcher attempted to focus on reading comprehension skill in this study . Reading comprehension skill was chosen by the researcher due to two reasons including. There has been much research on reading comprehension and linguistics through integration of some fields such as neuroscience, cognition, and education (Aldahhan, 2017). Neuroscience is the study of the brain and the nervous system in human beings. The electrical activity of the brain is caused by billions of nerve cells or neurons. There have been much research on reading comprehension and linguistics through

neuroscience(Aldahhan, 2017). Neurolinguistics is one of the subcategories of neuroscience. It is the branch of linguistics dealing with the relationship between language and functioning of the brain. Understanding the confidential points about how the brain functions when it deals with a text to read is an upward step in neuroscience (Qizi, 2020). Event-related potentials (ERPs) is one of the methods of brain activity measuring in neuroscience (Nidal&Malik, 2014), which has been used in this study in order to show how the brain functions when dealing with the reading comprehension section of the IELTS.(ERP) is an extraction of Electroencephalogram (EEG), but it is a task-related tool for measuring the electrical brain activities associated with a stimulus event (Bashashati , 2006; Ward, 2015).

One of the fundamental features of ERP components is that it allows us observe reading-related brain processes and it allows neural patterns to be documented at the millisecond level, at the same time that reading processes happens (Friederici, 2004; Steinhauer&Kasparian, 2020). Using ERP in language research helps to observe this hidden processing when there is no overt behavioral response (Andreassi, 2013).

As ERP can probe into the hidden processes involved in brain activity, the necessity of having paper-version tests is decreased (Steinhauer&Kasparian, 2020). Test takers in paper-version tests may use metacognitive strategies in order to pass the tests and not comprehensively comprehend the text (Qizi, 2020).

Research shows that the paper-version tests "did not very carefully predict candidate's proficiency level of

reading comprehension" (Hyatt & Brook, 2009, p.6), so revalidating it is an essential issue. For example, prior research on paper-version tests showed some distracter components such as "luck, guessing, anxiety levels, the speed of decision making and test taking capacity" (Kovaleno,2018, p.3) which offer a limitation for predicting and making differentiation amongtest takers' levels of reading comprehension (Kane &Bejar, 2014, cited in Kovalenko, 2018).Also, in the Corona virus emergence era the importance of computer-version tests is more felt. Due to the school and institutes closure there should be a good and valid platform in order to assess or check the learner's comprehension.

There are several researches that measure item discrimination of reading exam by using traditional ways (paper-version tests). The main problem of the paper-version tests is that they may not predict Meta cognitive strategies for answering reading comprehension questions (Kappenman& Luck,2011); thus, computer based tests based on ERP can be an alternative method to the existing method and One of the significance of using ERP method as a test during Corona virus era is that it is an economical and time saving method for studying the brain functions directly. Besides, this method has better time resolution (Abootalebi et al.,2006; Allent et.al,1992; Farewell& Donchin,1991; Rosenfeld, 2002). Since brain activities happen in millisecond, the ERP technique can show the moment-to-moment changes of the brain (Ortu, 2012).

ERP components can directly show item discrimination of reading comprehension questions. Moreover, this study employed ERP as an under-researched interdisciplinary area in examining computer-based reading comprehension.

Tests have been used to show the proficiency levels of the test takers and at the same time they are criterion and standard for educational decisions (Phakiti&Roever , 2011 , cited in Danuwijaya, 2018). With the increasing demands ofonline or computer-based tests due to the emergence of corona virus and school closure, it is necessary to decode which type of testing would be a better way of assessing. This study aims at choosing the best way of testing reading comprehension by using ERP components which allow researchers to observe the hidden processes behind the action of comprehending and the challenges which test takers deal with while making decisions about answering reading comprehension question.

Thus, tests should be a valid tool and predict each student's reading competence very accurately and distinguish students fairly. In this study one null hypothesis was developed: RQ1: Is there any significant difference between the paper-based and computer-version (ERP) of a same test?

Literature Review

The Neuroscience theory was selected to form the theoretical framework of this study since it covered the objectives of this study. The modern neuroscience allows scholars to study functions and malfunctions of the nervous system and how it can be changed (Brouwer& Crocker, 2017; Lorusso, et al., 2018; Small, et al., 2011). Neurobiology of language is one of the categories of neuroscience and it enables brain and language researchers to discover more about nerves system and its relationship with the language acquisition by combining different fields such as cognitive psychology, neurobiology, and

psycholinguistics (Schoffelen,et.al.,2019)The neurobiology of language reveals the neural activities underlying language processes (Small, et al., 2011, cited in (Bornkessel-Schlesewsky, I &Schlesewsky, M, 2019). Neurobiology of language focuses on production, comprehension, and acquisition of language communication disorders from a cognitive neuroscience perspective (Bear et al., 2016). In this study neurobiology of language was chosen to fit the objectives of the study.Nuerobiologyof language offers many valuable methods of language acquisition and learning to the linguists, teachers, and test designers (Roberts et al., 2018).

Emergence of the new brain imaging technologiessuch as Magnetic Resonance Imaging (MRI), Electroen cephalography technologies (EEG), and Event-related potentials (ERP) considerably helped the field (Nidal&Malik, 2014). EEG is a method in which electrical activities of the brain are recorded and measured (Budzynski t,& Evans, 2009).ERPs have central role in the field of neuroscience and language-related research (Bornkessel-Schlesewsky, I &Schlesewsky, M, 2019) and it led to the emergence of the neurobiology of language field (Small, 2008; Small, et al., 2011, cited in Bornkessel-Schlesewsky, I &Schlesewsky, M, 2019). ERPs are elicited due to the post-synaptic neural activities in the brain (Brouwer& Crocker, 2017).

Related Studies

Various research studies have been done in order to investigate the concept of ERP components and their effects on testing.ERP components can show the reaction of the brain to the semantic/ syntactic violation in the

language (Steinhauer et.al,2020).There are two popular components correlate with language, including the P600 andN400.

The P600 reflects late syntactic reanalysis" which in this study is the issue to be considered. Although there have been many findings from P600 about semantic and pragmatic anomalies, P600reflects the syntactic information most of the time (Courteau et al.,2019). Also, N400 is sensitive to the organization of semantic knowledge.

For instance, Tabullo et al.(2020) investigated the association between ERP components and reading comprehension abilities. To conduct the research, 24 students' reading comprehension ability was assessed with a cloze task. The cloze task consisted of strongly and weakly constraining sentences ending with an expected or unexpected final word. The results showed that the brain waves of students with better comprehension had smaller N400 for expected words and larger posterior positivity for unexpected endings. The conclusion suggested that the reading comprehension scores of students correlated with their ERPs. It means students who were better comprehenders can take advantage of predicting post lexical untis in sentences. Thus, "The N400 components" amplitude is standardly reduced for predictable words" (Szewczyk&Schriefers, 2018, p.6).

Bell, Angwin, Arnott, and Wilson (2019) in a research attempted to examine how children with hearing loss, using cochlear implants (CIS), process lexical-semantic incongruence. In order to conduct the research 12 children with CIS participated in the task while event-related

potentials were recorded. The task was spoken word-picture matching. Also 30 children participated in the research as typically hearing (TH) group. The results suggested that an N400 effect was elicited in both CIS and TH groups. The conclusion was the significant ability of CIS children in processing lexical-semantic incongruence and CIS group did not differ significantly from the TH group.

Recently, some scholars including Tromp (2018) have designed experiments in which mechanisms of comprehending the everyday language were under supervision completely controlled. In this research, participants were immersed in a virtual restaurant while they were equipped with EEG tools. During the presence of the participants in the restaurant some linguistic mismatches occurred intentionally by the guests who were present at the environment of the test. The researcher observed a reliable N400 effect due to the mismatch (Tromp et al.,&Hagoort, 2018). Thus, the N400 can predict semantic violations or mismatches.

A recent study by Zygouris et al. (2018) suggested the usages of ERP components in strengthening the language learning ability in children with dyslexia.

"Developmental dyslexia is defined as an unexpected specific and persistent failure to acquire efficient reading skills despite conventional instruction, adequate intelligence and socio cultural opportunity"(Zygouris et al., 2018, p.3).12 dyslexic children were asked tom participate in this research. They were divided into two groups; control and dyslexic group. The dyslexic group was under a 4-month instructional program which aimed at improving

the students' visual discrimination ability and improving their text comprehension. After the instructional and remediation program, the dyslexic group did not differ significantly from the control group. The results evidenced that both groups had shown P600 in their ERPs during the time they were answering the 8 neuropsychological tests. The result of this research suggested that dyslexic children can empower their linguistic abilities through a remediation program. Also the results provided an insight into how teachers of remediation programs can predict and evaluate their dyslexic students who want to improve their comprehension level or linguistic skills over time, and this prediction/evaluation process require a tool (ERP) to investigate the learners learning or comprehension processes or progress. As a result, language related components are the best tools for determining the discrimination of different sentences with violations or without.

There has been another effective research on the field of neurolinguistics. Martin et al.,(2016) investigated the difficulties experienced during sentence comprehension in a foreign language (L2). The participants were late learners of English and their first language was Spanish. The participants were exposed to three types of English sentences: 1.correct 2.semantic violation 3. Word knowledge violation. The sentences were semantically acceptable but factually untrue. While the participants were exposed to the tasks, their ERPs were recorded. The results showed that the N400 was elicited due to the semantic and mostly world knowledge violations. Thus, l2 readers were unable to have quick access to semantic violations than world knowledge violations (Martin et al., 2015).

A research by Weber-Fox and Neville (1996) investigated some evidence in order to support critical period (CP) in Chinese participants who had learned English. The researchers tested some Chinese participants with different levels of English proficiency because they had learned English at different ages. The researchers found some evidence for critical period (CP) in syntactic processing. Syntactic anomalies elicited P600 in participants' ERPs. The study concluded that even short delays in age of exposure to English may cause learners to be unable to analyze the syntactic structures of a language in the same manner as its native speakers do (Weber-Fox &Neville, 1996, cited in Steinhauer& Connolly, 2008).

Methodology
Participants and Setting

ERPs can shed light on the language teaching and testing fields specifically in the outbreak of Corona virus because in the Corona virus emergence era the importance of computer-version tests is more felt. Due to the school and institutes closure there should be a good and valid platform in order to assess or check the learner's comprehension.These developments are likely to provide significant benefits for the educational society and have implications for a diverse range of education, test designing, further research on reading comprehension

In order to collect the required data in this study, 10 Iranian IELTS candidates with an average age of 29 to 35 were selected. Participants included both genders (3 female, 7 male). Entry criterion for the participants was having an IELTS band score of 6-8. Also,the participants' agreement to participate in the study through a signed

written consent form was ensured. Furthermore, the participants filled out a written personal form in order to submit their normal and healthy neurological status. Participants were paid per session. The sampling procedure included purposive sampling. The setup of conducting this research was at the biomedical engineering laboratory of Islamic Azad University of Mashhad. Authorization from the ethics committee at biomedical engineering laboratory at Islamic Azad university of Mashhad to conduct this research was received. The following table demonstrates the demographic information of the participant's age, gender, educational background, reading score and IELTS band score. The reading score of participants is ranked from 15 to 20 that shows the spectrum of intermediate to advance students.

Due to the restricting criterion for participating in this research (having the IELTS band score of 6 to 8) the number of participants was10 initially. However, due to the time consuming process (around 3 hours) of the present research the number of participants decreased reliable data from 10 to9- 5 in each section of the test due to the presence or absence of artifacts or noises.Also, There has been much research in ERP field which were conducted with few participants, for instance, a study by Shahsavar et al. (2018) was conducted with 7 participants, Ghoshuni et al. (2013) with 10 participants. Liu, et al.,(2019) with 8 participants .Thus, it can be concluded that a few number of participants may not ruin the result of ERP based studies. Moreover, there was research by Hekmatmanesh, et al.,(2019) which was conducted by 9 participants. Another limitation is that gender of test takers could not be controlled.   The researcher also delimited the study in

different aspects. The age range of the participants was decided to be between 26 and 36 years old and participants had to be in healthy and normal neurological status. Another delimitation of this study was the participants' band score of the IELTS exam. Participants with the band score of 6-8 were allowed to participate in this experiment, because ERP is better extracted from the average and upper intermediate candidates in order to show the item discrimination of reading comprehension items in the IELTS exam. Also, in the current study, the researcher only considered the stem and categories of the reading comprehension questions and the researcher did not investigate the format of the reading comprehension questions.

Table 3.1 Demographic information of the participants

PARTICIPANTS	AGE	Reading score	GENDER	IELTS BANDSCORE
1	31	17	Male	7.0
2	33	20	Male	8.0
3	30	15	Female	6.5
4	32	17.5	Male	7.5
5	32	17	Male	7.0
6	29	17.5	Male	7.5
7	35	17	Female	7.0
8	29	17	Male	7.0
9	32	17	Male	7.0
10	31	16	Female	6.0

Instrumentation

This section includes the description of the instruments used in this study:

The Retired Version of the IELTS Exam

A retired version of the IELTS exam that was published in 2018 was used as stimuli. To conduct this, three academic reading passages were extracted from this retired version. Each included different number of reading comprehension questions. For instance, reading passage one had one section that contained 6 true-false items and , Reading passage two consisted of two parts; the first part contained 6 choosing the correct heading for each paragraph items and next part contained 4 matching names with ideas in the text. Reading passage three consisted of three sections. The first part contains 5 multiple choice that fit into "identify the underlying concept" part. The second part contained 6 open-ended items with possible fillers which fit into the "drawing logical inferences" part. The third part contained 3 true-false items which fit into the "following instruction" section. Some of the items in each category were omitted because more items could lead to fatigue which could have negatively affected the participants' brainwaves. Fatigue causes malfunction on ERPs (Boksem et al., 2005). Thus, some items were removed because at the end of the analysis their single trial ERP average could be calculated as zero and their artifact which shows the noise was ten. Thus, the final number of the items was 19. Eliminating the data with artifacts is used for filtering the ERPs (Mannan et al., 2016) and in order to have more reliable ERP results, data with high artifacts should be deleted (Salo et al., 2019).

ERP Recording Software

ERP is considered as a proper method to assess the brain performance among the other biomarkers for examining the effects of non-invasive brain stimuli (Shahsavar et al., 2018). This method is very suitable for this research based on the goal of the present study. By using EEG/ERP system, the participant's mental activity during the reading comprehension task was recorded. To carry out this, Psytask software was employed for showing the stimuli. Psytask software enable the researchers to design various psychological tasks (Aleksandrov et.al, 2019). In this study reading comprehension questions were used as the stimuli. Psytask has some potentials such as visual and auditory stimuli presentation, and reaction time measurement using patient event bottons (Pavlov & Tracey, 2017). It works simultaneously with WinEEG software and provides a synchronous stimuli presentation with EEG recording (Markovska-Simoska et al., 2018). The recorded data is used for extracting event related potentials (ERPs). The first step was adding the new task to the Psytask software with pictures of stimuli and then organizing the presentation order of the stimuli. Finally, the stimuli were randomized. The stimulus was in the second section of reading passage two in the reading comprehension test (Sample Selected items are available in the appendix A).Each stimulus was presented 10 times with duration of 2500ms by Psytask software. Additionally, stimuli were randomized. Finally, the researcher analyzed and interpreted the data by WINEEG and MATLAB software.

WINEEG is a software that allows the researcher to record, edit, and analyze the participants' brainwaves (www. Mitsar-medical.com).WINEEG recording is a

useful brain imaging technique and is a very economical and trustworthy way to record the brain activities (www.Mitsar-medical.com).

MATLAB is a numerical computing system and it integrates computation, data analysis, engineering graphics, and algorithm developments (Cruz, 2016). It hasbeen used in different fields such as engineering, science, and economics (www. Mitsar-medical.com).

Procedure

Initially, the participant's agreement to participate in this study was obtained through a signed written consent form. Participants' IELTS band score reports were obtained(Two Sample report cards are in the Appendix B) through the paper-based IELTS reading test. The procedure of this research was divided into two phases:

Phase One: Reading Comprehension

At first, the stages of the research were explained to each participant separately. Second, three Academic Reading passages that were extracted from the retired version of IELTS exam were given to the participants as a paper-based test and then they were asked to read the passages and answer the reading comprehension questions on a paper answer sheet. According to the IELTS exam standards, the time given to each passage was 20 minutes and each participant had to spend 20 minutes on each passage. The first reading passage consisted of 6 true-false and not given items, the second one included 6 items, 4 of which were omitted as they related to the choosing headline for each paragraph, also there were 4 questions of underlying concept in the format of matching names with headlines that one of which was omitted, and the third

passage consisted of 3 sections, the first section included 5 multiple choice items,two of which were omitted. The next section of the third passage had 6 matching items that 4 of which were omitted, and the third part of the third passage had 3 true-false items. Totally, there were 6 sections with 19 items. The participants' reading comprehension scores were calculated.

Phase Two: ERP Recording

The answer sheets of the previous phase were taken from the participants, but reading passages were still available to them. In order to record the EEG signals, 19 electrodes (O1, O2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T5, P3, Pz, P4, T6, Fp1, Fp2) were mounted in an elastic cap on the scalp of the participants according to the international 10-20 system (stimuli were presented in ODDBALL paradigm)Studies of cognition often use an oddball paradigm to study effects of stimulus novelty and significance on information processing (Pavlov & Tracey, 2017).To conduct the experiment, the skin surface was cleaned using the gel and then the attachment of the electrodes was done. Electro-Gel was used for decreasing the impedance of the skin surface into 10'Ω in order to increase the signals. The electrode impedances were also kept below 10 k'Ω. Initially, the EEG signal was recorded from 19 electrodes (O1, O2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T5, P3, Pz, P4, T6, Fp1, Fp2) attached to the scalp of the participants and both earlobes were connected, but finally the researcher decided to just consider the Fz, Cz, and Pz positions according to the international 10-20 system as some scholars including Shahsavar et al. (2018) conducted their experiment based on the Fz, Cz, and Pz positions. A 15-inch wide screen monitor was used for the

visual stimuli. To show the stimuli the Psytask software was employed. 19 reading comprehension questions were shown separately and their answers were shown randomly and the participants were allowed to choose just one answer for each reading comprehension question.The process of choosing the answers was different from the pen-paper process in the previous phase. The researcher requested the participants to left click as soon as they see the right answer. Each item was shown 10 times randomly according to the ODDBALL paradigm. While the participants were doing the tasks; simultaneously, their EEG signals were being recorded. Finally, ERP signals were analyzed by MATLAB and WINEEG software in order to identify the item discrimination of reading comprehension questions.

Study Design and Analysis

   The method of this study is quantitative. In this model the researcher collected data in two phases in which qualitative data was collected first; then, qualitative data (polarity and brain waves) were changed to quantitative data by employing independent T-test. The qualitative data changed to quantitative because the data should have been more tangible and touchable research (Creswell &Clark ,2017). Finally, data was interpreted. In the present study, in the first phase, participants were supposed to read and answer the reading comprehension section of IELTS exam on a paper answer sheet. In the second phase of the study, researcher recorded the brain waves of each participant by using PSYTASK software for showing the stimuli and WINEEG software for recording the participant's brain waves. Using the PSYTASK and WINEEG software were

simultaneous. By using WINEEG , the researcher extracted ERP graphs from brain waves. Then, the researcher changed the quantitative data (ERP graphs) to qualitative data (numbers). For analyzing and interpreting the data, MATLAB software was used. Also, One-Way ANOVA statistical test was employed in order to compare the mean scores and to find the potential difference between Paper-Based tests and computer-Based tests.

Preliminary Analysis

In order to collect the data, participants participated in an EEG recording session through which their brain waves were recorded. The electrode impedances were also kept below 10 kΏ. Figure 4.1 displays one of the participant's EEG signals which were recorded during the presentation of the stimuli. Although EEG signals contain rich information, they are very challenging and cannot present the meaningful information clearly. Thus, ERPs should be extracted from EEG signals in order to illustrate more exact interpretations (Scheffler&Senturk, 2019). Thus, in this study recorded EEG signals were filtered through analysis and computing ERPs by WINEEG software. In order to extract ERPs from EEG signals, the parameters for ERP computation including artifact processing level, thresholds for channels, and synchronization type were determined in the WINEEG software.

Then, the reliable data format was changed to txt files and renamed in order to import to MATLAB software. The reliable data was imported to the work space of MATLAB software and matrixes (size=250) were made for each participant.

Figure 4.1 one of the participant's EEG signals
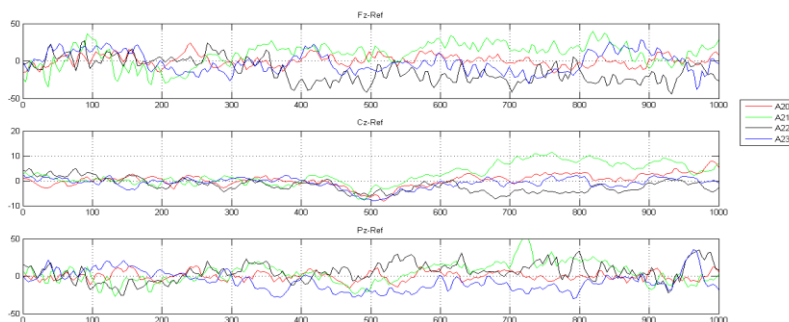
The size of each matrix should be the same in order to have homogenous data analysis. In the next level, ERP graphs for each item in each section of reading comprehension part were extracted from MATLAB software. The ERP analysis focused on two time windows: the N400 component (300-500 m.s) and the P600 component (500-800 m.s). The ERPs from Fz, Cz, and Pz channels are elicited from each item of the second section of reading comprehension passage two. Four ERPs that related to 4 items of the reading section 2 are illustrated in three channels and the amplitude of A20 is different from other items in different channels. . The sum and difference of target items from non target items in each section was also calculated for N400 and P600 components. Moreover, an Anova test was run to investigate the significant differences between reading comprehension sections regarding N400 and P600 components. Finally, bar graphs for each two sections were extracted.

Results

In order to check the reliability of the both paper and computer-based tests, the internal consistency of the tests was calculated and high-reliability coefficients (for PBT, $\alpha= 94$ & for CBT, $\alpha= 95$) were achieved. Results showed

that the highest mean score was found in CBT, with a relatively higher mean score by .53 points. Test takers' mean score on CBT (M = 46.54, SD=17.63) was a little bit higher than their mean score on the PBT (M=46.11, SD=13.3). Also, the standard deviation in CBT was higher than in PBT. It meant that the dispersion of scores from the mean score in CBT was higher than in PBT; consequently, it was concluded that Standard Error of Measurement (SEM) in CBT was lower than in PBT.

According to the findings of the One-Way ANOVA test (Table 4.1), there was not any statistically significant difference in scores between PBT and CBT at a .05 level. Based on the results of the score analysis of two testing sessions, the Sig. value was .896 at P≤0.05. This amount of significance value at 119 (N-1) degree of freedom in a .05 level revealed that there was no significant difference between two sets of scores obtained from both paper-based and computer-based tests and the test scores of participants were not different in paper-based and computer-based versions of the test (Sig=.896, P>0.05).

Table 4.1 One-way ANOVA comparing scores of participants in paper and computer- based tests

ANOVA

	Sum of Squares	D.F.	Mean Square	F	Sig.
Between Groups	4.269	1	4.269	0.17	.896
Within Groups	14236.123	235	247.306		
Total	14240.392	236			

The present study aimed at showing thedifference between paper-based and computer-based tests in the Corona virus era and it revealed the significance of using

direct methods such as ERP rather than indirect testing methods such as paper based tests. Findings showed that although participants may use Metacognitive strategies to answer the tests in paper-based test, they have the same results in computer-based tests. This study have many implications for teachers, institutes, and specifically researchers.

The first groups who can benefit the results of this study are book writers, material designers, and generally test designers who attempt to provide more valid tests. By using the results of the current study, the test designers can design placement tests or other types of tests which are categorized from easy to difficult.

Moreover, the second group who can directly benefit the results of the current research is the IELTS designers' team. Test designers,policy makers in educational systems, and teachers who attempt to provide more validtest specifically reading comprehension items with determined difficulty level can take advantage of the results of this research. Thus, IELTS designers can use the results of this study to have well designed items and to promote the validity of the test. Well-designed items are used to improve an understanding of a test.

As for language teachers, they are recommended to check the item discrimination of the tests through neuroscience to have better understanding of the questionsdifficulty level. They can adjust the difficulty level of the items in tests with the proficiency level of their students. For instance, when the teachers have to take testsfrom the advanced level students. Also, they can accord theirtesting with the comprehension level of their

students. They can set their tests based on ERP evidence in order to test effectively. The results of this study are used to suggest possible enhancements to the teachers and the students.

The results of this research also have implications for teachers specially ones who test IELTS reading comprehension section. Teachers can plan their instructional activity involved in teaching reading comprehension skill based on the results of the current study. They can benefit using ERP recording technology results to know their students' reading comprehension level better and make lesson plans accordingly to divide their students into homogeneous groups. Richard and Trowler (2018) suggests that students can learn better if they are in homogeneous groups with the same level of proficiency so teachers based on the result of the current research can use ERP components and distinguish their students' level of proficiency in reading directly. Moreover, Ding ,Chen, and Huo (2019) believe that there is no doubt that using new technologies in language teaching field has a central role and teachers can be researchers of their own classes.

Suggestion for Further Research

This study, as an interdisciplinary study, it can open new visions for other studies. There might be some possible gaps or limitations in the current study. Thus, the following suggestion is provided for further research. Due to the limited sample size in this study, replication of the current study can be done in order to increase the generalizability of the study findings. In other words, replication of this study with a larger sample size can provide findings that can be generalized to a bigger target population.

Reference

- Al Dahhan, N. Z., Kirby, J. R., Brien, D. C., & Munoz, D. P. (2017). Eye movements and articulations during a letter naming speed task: children with and without dyslexia. *Journal of Learning disabilities*, *50*(3), 275-285.https://doi.org/10.1177%2 F0022219415618502
- Alderson, J. (2000). *Assessing Reading*. Cambridge UK: Cambridge University Press.
- Aleksandrov, A. A., Dmitrieva, E. S., Volnova, A. B., Knyazeva, V. M., Gainetdinov, R. R., &Polyakova, N. V. (2019). Effect of trace amine-associated receptor 1 agonist RO5263397 on sensory gating in mice. *NeuroReport*, *30*(15), 1004-1007.
- Andreassi, J. L. (2013). *Psychophysiology: Human behavior & physiological response*. Psychology Press.
- Bashashati, A., Noureddin, B., Ward, R. K., Lawrence, P. D., & Birch, G. E. (2006). An experimental study to investigate the effects of a motion tracking electromagnetic sensor during EEG data acquisition. *IEEE transactions on biomedical engineering*, *53*(3), 559-563.
- Bear, M. F., Connors, B. W., & Paradiso, M. A. (2018). *Neurowissenschaften:EingrundlegendesLehrbuchfürBiologie, Medizin und Psychologie*. Springer-Verlag.
- Bell, N., Angwin, A. J., Wilson, W. J., & Arnott, W. L. (2019). Reading development in children with cochlear implants who communicate via spoken language: a psycholinguistic investigation. *Journal of Speech, Language, and Hearing Research*, *62*(2), 456-469.
- Boksem, M. A., Meijman, T. F., &Lorist, M. M. (2005). Effects of mental fatigue on attention: an ERP study. *Cognitive brain research*, *25*(1), 107-116.
- Bornkessel-Schlesewsky, I., &Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in psychology*, *10*, 298.
- Brouwer, H., & Crocker, M. W. (2017). On the proper treatment of the N400 and P600 in language comprehension. *Frontiers in Psychology*, *8*, 1327.

- Budzynski, T. H., Budzynski, H. K., Evans, J. R., &Abarbanel, A. (Eds.). (2009). *Introduction to quantitative EEG and neurofeedback: Advanced theory and applications*. Academic Press.
- Courteau, E., Martignetti, L., Royle, p., &Steinhauer, K. (2019). Eliciting ERP Components for Morphosyntactic Agreement Mismatches in Perfectly Grammatical Sentences. *Frontiers in psychology, 10,* 1152.
- Cruz, C. D. (2016). Genes Software-extended and integrated with the R, Matlab and Selegen. *ActaScientiarum. Agronomy*, *38*(4), 547-552.
- Danuwijaya, A. A. (2018). Item analysis of reading comprehension test for post-graduate students. *English Review: Journal of English Education*, *7*(1), 29-40.
- Ding, H., Chen, K., &Huo, Q. (2019). Compressing CNN-DBLSTM models for OCR with teacher-student learning and Tucker decomposition. *Pattern Recognition*, *96*, 106957.
- Friederici, A. D. (2004). Event-related brain potential studies in language. *Current neurology and neuroscience reports*, *4*(6), 466-470.
- Ghoshuni, M., Firoozabadi, M., Khalilzadeh, M. A., &HashemiGolpayegani, M. R. (2013). Variation of wavelet entropy in electroencephalogram signal during neurofeedback training. *Complexity*, *18*(3), 18-23.
- Hasson, U., Egidi, G., Marelli, M., & Willems, R. M. (2018). Grounding the neurobiology of language in first principles: The necessity of non-language-centric explanations for language comprehension. *Cognition*, *180*, 135-157.
- Hekmatmanesh, A., Wu, H., Motie-Nasrabadi, A., Li, M., &Handroos, H. (2019). Combination of discrete wavelet packet transform with detrended fluctuation analysis using customized mother wavelet with the aim of an imagery-motor control interface for an exoskeleton. *Multimedia Tools and Applications*, *78*(21), 30503-30522.
- Hyatt, D., & Brooks, G. (2009). Investigating Stakeholders' Perceptions of IELTS as an Entry Requirement for Higher Education in the UK. *International English Language Testing System (IELTS) Research Reports 2009;10(1)*.

- Isaacs,   T.,&Trofimovich,   p.   (2012).   Deconstructing Comprehensibility: Identifying the Linguistic Influences on Listeners' L2 Comprehensibility Ratings. *Cambridge University Press 34(3).*
- Kappenman, E. S., & Luck, S. J. (2011). of Brainwave Recordings. *The Oxford handbook of event-related potential components*, 3.
- Kovalenko, M. (2018). The Validation Process in the IELTS Reading Component: Reading Requirements for Preparing International Students. *Journal of Language and Education*, 4(1),63-78. Doi:10.17323/2411-7390-2018-4-1-63-78.
- Liu, X., He, P., Chen, W., &Gao, J. (2019). Improving multi-task deep neural networks via knowledge distillation for natural language understanding. *arXiv preprint arXiv:1904.09482*.
- Lorusso, L., Piccolino, M., Motta, S., Gasparello, A., Barbara, J. G., Bossi-Régnier, L., ... &Molnár, Z. (2018). Neuroscience without borders: preserving the history of neuroscience. *European Journal of Neuroscience*, *48*(5), 2099-2109.
- Mannan, M. M. N., Kim, S., Jeong, M. Y., & Kamran, M. A. (2016). Hybrid EEG—Eye tracker: Automatic identification and removal of eye movement and blink artifacts from electroencephalographic signal. *Sensors*, *16*(2), 241.
- Markovska-Simoska, S., Pop-Jordanova, N., & Pop-Jordanov, J. (2018). Inter-and Intra-Hemispheric EEG Coherence Study in Adults with Neuropsychiatric Disorders. *prilozi*, *39*(2-3), 5-19.
- Martin, C. D., Garcia, X., Breton, A., Thierry, G., & Costa, A. (2016). World knowledge integration during second language comprehension. *Language, Cognition and Neuroscience*, *31*(2), 206-216.
- Moore, T., Morton, J., & Price, S. (2012). Construct validity in the IELTS academic reading test: A comparison of reading requirements in IELTS test items and in university study. In *Studies in language testing* (pp. 120-211). Cambridge University Press Cambridge.
- Nidal, K., & Malik, A. S. (Eds.). (2014). *EEG/ERP analysis: methods and applications*. NY: Crc Press.

- O'Reilly, M., & Lester, J. N. (2017). *Examining mental health through social constructionism: The language of mental health*. USA: Springer.
- Pavlov, V. A., & Tracey, K. J. (2017). Neural regulation of immunity: molecular mechanisms and clinical translation. *Nature neuroscience*, *20*(2), 156.
- Qizi, M. F. S. (2020). IMPLEMENTATION OF NEURO-LINGUISTICS PROGRAMMING IN COMMUNICATIVE TEACHING OF ENGLISH. *Проблемыпедагогики*, (1 (46)).
- Rastelli, S. (2018). Neurolinguistics and second language teaching: A view from the crossroads. *Second Language Research, 34*(1), 103-123.
- Roberts, L., Gonzalez Alonso, J., Pliatsikas, C., & Rothman, J. (2018). Evidence from neurolinguistic methodologies: Can it actually inform linguistic/language acquisition theories and translate to evidence-based applications?. *Second Language Research*, *34*(1), 125-143.
- Salo, K. S. T., Mutanen, T. P., Vaalto, S. M., &Ilmoniemi, R. J. (2020). EEG artifact removal in TMS studies of cortical speech areas. *Brain Topography*, *33*(1), 1-9.
- Schoffelen, J. M., Oostenveld, R., Lam, N. H., Uddén, J., Hultén, A., &Hagoort, P. (2019). A 204-subject multimodal neuroimaging dataset to study language processing. *Scientific data*, *6*(1), 1-13.
- Shahsavar, Y., Ghoshuni, M., &Talaei, A. (2018). Quantifying clinical improvements in patients with depression under the treatment of transcranial direct current stimulation using event related potentials. *Australasian physical & engineering sciences in medicine*, *41*(4), 973-983.
- Small, S. A., Schobel, S. A., Buxton, R. B., Witter, M. P., & Barnes, C. A. (2011). A pathophysiological framework of hippocampal dysfunction in ageing and disease. *Nature Reviews Neuroscience*, *12*(10), 585-601.
- Steinhauer, K., Connolly, J. F., Stemmer, B., & Whitaker, H. A. (2008). Event-related potentials in the study of language. *Concise encyclopedia of brain and language*, 91-104.
- Steinhauer, K., & Kasparian, K. (2020). Brain Plasticity in Adulthood—ERP Evidence for L1-attrition in Lexicon and

MorphosyntaxAfter Predominant L2 Use. *Language Learning*, *70*(S2), 171-193.
- Szewczyk, J. M., &Schriefers, H. (2018). The N400 as an index of lexical preactivation and its implications for prediction in language comprehension. *Language, Cognition and Neuroscience*, *33*(6), 665-686.
- Tabullo, Á. J., Shalom, D., Sevilla, Y., Gattei, C. A., París, L., &Wainselboim, A. (2020). Reading Comprehension and Predictability Effects on Sentence Processing: An Event-Related Potential Study. *Mind, Brain, and Education*, *14*(1), 32-50.
- Tromp, J., Peeters, D., Meyer, A. S., &Hagoort, P. (2018). The combined use of virtual reality and EEG to study language processing in naturalistic environments. *Behavior Research Methods*, *50*(2), 862-869.
- Urquhart, A. H., & Weir, C. J. (2014). *Reading in a second language: Process, product and practice*. Routledge.
- Ward, J. (2015). *The student's guide to cognitive neuroscience*. UK: Psychology Press.
- Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of cognitive neuroscience*, *8*(3), 231-256.
- Zygouris, N. C., Avramidis, E., Karapetsas, A. V., &Stamoulis, G. I. (2018). Differences in dyslexic students before and after a remediation program: A clinical neuropsychological and event related potential study. *Applied Neuropsychology: Child*, *7*(3), 235-244.

✱✱✱