British Journal of Education, Society & Behavioural Science
6(3): 196-209, 2015, Article no.BJESBS.2015.056
ISSN: 2278-0998

SCIENCEDOMAIN *international*
www.sciencedomain.org

SCIENCEDOMAIN

Reading Comprehension Testing Techniques and Cognitive Fatigue: A Study on Construct Irrelevant Factors

Mojgan Rashtchi[1*], Mona Abbaszadeh[1] and Narges Tabarraee[1]

[1]Department of TEFL, Faculty of Foreign Languages, Islamic Azad University, North Tehran Branch, Iran.

Authors' contributions

This work was carried out in collaboration between all authors. Author MR designed the study, wrote the protocol and supervised the work. Authors MA and NT carried out all laboratories work and performed the statistical analysis. Author MR managed the analyses of the study. Author MA wrote the first draft of the manuscript. Author MR managed the literature searches and edited the manuscript. All authors read and approved the final manuscript.

Article Information

DOI:10.9734/BJESBS/2015/14976
Editor(s):
(1) Timothy Rasinski, Kent State University, USA.
(2) Ali Kazemi, Department of English, Yasouj University, Iran.
(3) Rajendra Badgaiyan, Neuromodulation Scholar, University of Minnesota, Minneapolis, MN, USA.
Reviewers:
(1) Adriana Marques de Oliveira, Faculdade de Filosofia e Ciências – FFC/UNESP (Marília/SP/Brasil), School of Science and Philosophy of São Paulo State University (Unesp) / FFC/ UNESP/Marília (SP).
(2) Anonymous, Turkey.
Complete Peer review History: http://www.sciencedomain.org/review-history.php?iid=820&id=21&aid=7756

Original Research Article

Received 30th October 2014
Accepted 29th December 2014
Published 10th January 2015

ABSTRACT

The aim of the present study was to investigate the impact of different reading comprehension testing techniques on the arousal of cognitive fatigue in test takers. Additionally, the researchers investigated whether cognitive fatigue might be affected by the test takers' IQ and cognitive styles. Raven's Advanced Progressive Matrices and Group Embedded Figures Test were used for measuring test takers' IQ and cognitive style, respectively. Two groups of 15 students were asked to answer the MC passage comprehension and MC rational cloze test in 90 minutes. Subsequently, both groups were asked to perform the Wisconsin Card Sorting Test and Continuous Performance Test used to measure cognitive fatigue. The findings showed no significant difference between the two reading groups in terms of their cognitive fatigue. Moreover, no significant differences were

**Corresponding author: E-mail: mojgan.rashtchi@gmail.com;*

found among the impact of testing technique, IQ, and cognitive style on the test takers' cognitive fatigue.

Keywords: Cognitive fatigue; passage comprehension; rational cloze; reading; testing techniques.

1. INTRODUCTION

Valid measures of reading comprehension require a comprehensive description of the processes involved in understanding a text together with the constructs related to the reliable variance. The measures also call for identification of measurement errors and the way they interfere with the reliable variance. Identifying reliable variance as well as measurement errors contribute to the reliability of tests. Reliability, as Bachman [1] argued, depends upon the ability to distinguish the effect of the abilities a test intends to measure from the effect of other factors. Some factors, as Bachman [1] maintained, have systematic effect on test performance like test method facets and test takers' attributes, while some have unsystematic effect such as administration environment, test takers' emotional status, and the like. Thus, any factor (except the ability the test intends to measure) which affects test takers' performance is a threat to the reliability of the test, and consequently to its validity.

Correspondingly, testing reading comprehension is an important issue because a variety of factors seem to affect testees' performance. A review of the literature indicates that many studies have aimed to show the sources of unsystematic variance in reading comprehension tests. Factors such as automatic recognition skills, world knowledge, strategic knowledge, and metacognitive knowledge [Ediger, as cited in 2], working memory capacity (Carretti, Borella, Cornoldi, & Beni, [3]; Daneman & Carpenter, [4]; Mih & Mih, [5]; Oakhill, Hartt, & Samols, [6]), anxiety (Sellers, [7]), and motivation (Morgan & Fuchs, [8]) signify the individual differences, which according to Alderson [9], cause test takers to perform differently.

Also, the effect of reading comprehension testing techniques on test performance (Hassani & Maasum, [10]; Keenan & Meenan, [11]; Kobayashi, [12]), mental process and strategy investment (Han, [13]; Kendeou, Papadopoulos &Spanoudis, [14]; Rupp, Ferne, & Choi, [15]), and their relationship with individual differences (Ozuru, Dempsey, & McNamara, [16]; Pishghadam & Tabatabaian, [17]) have been the subject of many studies. However, the impact of reading comprehension testing techniques on the arousal of cognitive fatigue during test performance, as review of the literature shows, has not been addressed previously.

According to Van der Linden, Frese, and Sonnentag [18], increase in cognitive fatigue coincides with changes in motivation, information processing, and mood. Moreover, cognitive fatigue results in decrease in task engagement and an increased resistance to allocate further effort to task accomplishment (Van der Linden et al., [18]). Therefore, the issue of cognitive fatigue in test taking especially in high-stakes tests seems to be an important factor overlooked by educational researchers. The aim of the present study was to examine test takers' cognitive fatigue after their performance on different reading comprehension tests.

2. REVIEW OF THE LITERATURE

One of the most commonly used techniques for measuring reading comprehension ability in high-stakes tests (TOFEL, IELTS, etc.) is using reading passages followed by a number of multiple-choice questions to measure test takers' comprehension. Developing such tests, according to Alderson and Bachman [19], is time-consuming and requires expertise on the part of the test-developers, because choosing plausible distractors which could discriminate between the weak and strong students is difficult. However, due to the objectivity of such tests and at the expense of validity, the technique is wildly applied throughout the world. A disadvantage of these traditional MC passage comprehension tests is that alternatives which are developed based on test constructors' personal interpretations of the texts may interfere between the passage writer and test takers (Jafarpur, [20]).

In addition to MC passage comprehension tests, research evidence showed that rational cloze test can be applied as a valid measure of reading comprehension ability (Abraham & Chapelle, [21]; Brown, [22]; Chapelle & Abraham, [23];

Greene, [24]; Gellert & Elbro, [25]; Han, [13]; Mckamey, [26]; Storey, [27]; Tajeddin, [28]; Yamashita, [29]). Also, short-answer (Alderson, [9]) and summary writing tests (Bensoussan & Kreindler, [30]) are considered to be valid instruments for measuring reading comprehension; however, due to the lack of practicality, they are not generally used in large-scale tests. Although none of the reading comprehension measures enjoys appropriate validity (Alderson & Bachman, [19]), these varieties of test techniques might have various impacts on test takers' performances, which could be due to their personal differences. Thus, different testing techniques can create measurement errors or test bias and jeopardize the reliability and validity of tests, and consequently harm test fairness considerations.

2.1 Construct Irrelevant Variances and Reading Measurement

The psychological and physiological characteristics like motivation, intelligence, personality, anxiety, gender, and cognitive style of test takers can contribute to unreliable variance. These individual differences can produce some systematic construct-irrelevant variances called test bias, leading to an invalid interpretation of test scores (McNamara & Roever, [31]). For example, in terms of gender differences, findings indicated that females tend to perform better on the same reading comprehension test than males (Chiu, Chow & McBride-Chang, [32]). The results of some studies signified that students with higher motivation to read are more inclined to become skilled readers than less motivated students (Morgan & Fuchs, [8]). Also, there is some evidence that more anxious students tend to recall less passage content than do the participants who experience less anxiety (Sellers, [7]).

The capacity of working memory, according to Daneman and Carpenter, [4], is also amongst individual factors which affect text comprehension. It is argued that skilled readers have higher working memory capacity than the less-skilled readers (Carretti, Borella, Cornoldi, & Beni, [3]; Daneman & Carpenter, [4]; Mih & Mih, [5]; Oakhill, Hartt, & Samols, [6]). Cognitive styles have been shown to affect reading comprehension, as well. Some studies indicated that field-dependent individuals perform differently on cloze and multiple choice tests than field-independent individuals (Alderson, [9]). Additionally, the study on the relationships between IQ and four different reading comprehension test formats (multiple choice, cloze test, c-test, summary writing) by Pishghadam and Tabataba'ian [17] showed no correlation between IQ and performance on MC tests (except for Arithmetic sub-section); however, performances on cloze tests, c-tests, and summary writing were shown to be related to IQ.

2.2 Testing Technique

In addition to psychological and physiological differences, "the characteristics of the method used to elicit test performance" (Bachman, [1], p. 111) are considered to impact language test performance (Alderson, [9]; Bachman, [1]; Bachman & Palmer, [33]). Therefore, test takers' performances on a test could not be merely attributed to their knowledge but to their ability to answer certain item formats (Baker, [34]). Some studies have investigated test method (or test technique) effect on language test performance. Kobayashi [12] found that text organization and test format have a significant impact on students' performance. His study revealed that when tests are clearly structured, more proficient students achieve better scores in summary writing and open-ended questions, while the structure of the test hardly affects the performance of less proficient students. He concluded that by considering the effect of test structure and format on students' performances, validity can increase.

Also, Kendeou, Papadopoulos, and Spanoudis [14] found that three different reading comprehension tests (passage comprehension, maze reading test- similar to cloze test, and recall test) impose different processing demands on young readers. Moreover, Pishghadam and Tabataba'ian [17] argued that since one single format might produce an element of bias, using a mixture of different test types would be more fruitful. Bachman [1] maintained that test performance which is affected by test method as well as the interaction of test method and individual characteristics may decrease the validity of tests, implying that it is important not only to identify but also to minimize these effects. However, provision of an in-depth analysis of "the issue of test method facet, and the fact that different testing techniques or formats may themselves test non-linguistic cognitive abilities or give rise to affective responses" seems to be a must in language testing (Alderson & Bachman, [19], p. 202).

2.3 Cognitive Fatigue

Van der Linden et al. [18] defined cognitive fatigue as "a psycho-physiological state resulting from sustained performance on cognitively demanding tasks" which coincides with "changes in motivation, information processing, and mood" (p. 484). They also stated that cognitive fatigue commonly results in a decrease in task engagement. Increase in fatigue, also, appears to be associated with increase in number of errors and decrease in efficiency (Grandjean, [35]). According to Van der Linden et al. [18], one of the important issues in studies related to fatigue is whether cognitive control of behavior changes under fatigue. They suggested that the decline in executive control depends on the type of task (demanding or undemanding), the amount of the required control (executed in a more or less automatic way), and the allotted time. Executive functioning which appeared to be a useful quality for detecting cognitive fatigue (Mizuno & Watanabe, [36]; Moller, [37]) refers to the ability to regulate perceptual and motor processes in order to respond in an adaptive way to novel or changing task demands (Baddeley & Logie, as cited in [18]).

Van der Linden et al., [18] reported that fatigued participants showed more performance deficits than non-fatigued subjects in performing tasks which required planning and flexibility, as well as generating and testing hypotheses. They maintained that working memory, which is responsible for the executive control, is used for the temporary storage, active monitoring, and manipulation of information. Thus, cognitive fatigue would affect performance on those types of tasks which require the involvement of working memory. All in all, cognitive fatigue seems to potentially affect test performance.

2.4 Purpose of the Study

This study aimed to investigate whether different reading comprehension testing techniques have any effect on producing cognitive fatigue. The researchers further attempted to verify whether the test takers' cognitive fatigue after the exam session was more affected by the different reading comprehension measurement techniques than their individual differences on test performance. In this study, IQ and cognitive style were analyzed as individual differences of the test takers. To rephrase, the purpose of the present study was to address cognitive fatigue as a by-product of different testing techniques and ultimately to alert test makers of other measurement errors which are required to be taken into account during test construction. Controlling the intrusion of cognitive fatigue in test taking process is particularly significant when test results deal with making decisions related to individuals' future life, work, career, and profession. If the purpose of testing, in general, is to tackle and measure the constructs which tests intend to measure, taking error factors (like cognitive fatigue) into consideration seems to be relevant. Altogether, the results of the study can contribute to reliability and validity, and especially to test fairness considerations. Following the aforementioned purposes, the present study attempted to answer the following research questions:

RQ1 Is there any difference in test takers' cognitive fatigue after their performance on passage comprehension and rational cloze?

RQ2 Do reading measurement techniques, IQ, and cognitive style have the same impact on the test takers' cognitive fatigue?

3. METHODS

3.1 Participants

Participants were initially comprised of 45 senior college students majoring in English at Islamic Azad University, North Tehran Branch, Iran. However, the researchers lost 15 participants due to the mortality factor. Thus, the number of the participants whose data were analyzed reduced to 30 (25 females and five males).

3.2 Instrumentation

The following instruments were employed in this study:

- The first instrument was an MC passage comprehension test. The test was comprised of six passages followed by 56 MC comprehension questions. Five passages were taken from the reading part of the paper-based TOEFL (2003) and one from the reading part of the paper-based TOEFL (1975). The readability of this test based on Flesch reading ease appeared to be 45.9. The reliability estimated through Cronbach's alpha showed a high index (r=.86). The time allotted to this test was 90 minutes.

- The second instrument was the MC rational cloze test which consisted of six passages with 60 gaps followed by 60 MC questions. Overall, 20 nouns, 23 verbs, 12 adjectives, 2 adverbs, 2 cohesive devices, and 1 preposition were deleted from the passages. The six passages were selected from the General English section of the University Entrance Examinations for TEFL MA program in Iran. The readability of this test based on Flesch reading ease appeared to be 45.8. The reliability index of the test estimated through Cronbach's alpha was acceptable (r= .77). Similar to the previous test, 90 minutes was allocated to answering this test.
- The third instrument was the free computerized version of Wisconsin Card Sorting Test (WCST), (PEBL Version 0.13). The WCST consists of four key cards and 128 response cards with geometric figures which vary in color, shape, and number. The participants were asked to discover the sorting rule through trial and error using the feedback provided by the software after each trial. Once the participants found out the rules, they had to maintain the rule for ten consecutive cards when the sorting rule changed without notice. Then again, they had to discover the new rule and sort the cards.
  This test is commonly applied to assess the executive control under fatigue (Nyhus, Barceló, [38]; Somsen, van der Molen, Jennings, & van Beek, [39]; van der Linden et al., [18]). In order to interpret participants' performances on this test, usually two or three scores are assessed including the numbers of perseverative errors, completed categories, and non-perseverative errors (Barceló & Knight, [40]; Greve, [41]; van der Linden et al., [18]). In this study, the first two indices (i.e. the number of perseverative errors and completed categories) were investigated as a sign of cognitive fatigue. Preservative errors indicate that people tend to continue applying previous sorting rules that are no longer valid (van der Linden et al., [18]).
- The fourth instrument was the free computerized version of Continuous Performance Test (CPT), (PEBL Version 0.13). This test is frequently used to obtain individuals' ability to sustain attention over time (Riccio, Reynolds, Lowe, & Moore, [42]). In this test, participants sit and watch the alphabet letters on a monitor during the Inter-Stimulus Intervals (ISI) of 1, 2, and 4 seconds. This test usually takes 15 minutes during which the participants have to press the space key on the keyboard as soon as they see the letters except for the letter X. In order to interpret participants' performances, usually two types of errors are investigated: commission errors and omission errors (Armengol, [43]; Kieling, Roman, Doyle, Hutz, & Rohde, [44]; Riccio, Reynolds, Lowe, & Moore, [42]). "Omission" errors indicate failure to respond to the target letters (non-Xs) and "commission" errors show that the responses are given to non-target letters (Xs) (Conner, [45]). The software (PEBL Version 0.13) provides report for each participant in the ISI of 1, 2, and 4 seconds and a total or "pooled" which is the aggregation of all errors participants make (the pooled sections were used in this study). In addition to these two types of errors, this study investigated the "correct reaction time mean score" provided in the software report which demonstrates the average speed of correct responses for the entire test.
- The next instrument was Group Embedded Figures Test (GEFT). This test is commonly applied to investigate test takers' cognitive style (Field-dependent [FD] versus field-independent [FID]). This test consists of three sections. The first section serves as a preparatory section for doing the next two sections, and consists of 18 complex figures each with an embedded simple figure and must be completed in 12 minutes. The number of correct responses is aggregated in order to find test takers' cognitive styles. Scores on GEFT may range from 0 (extreme FD) to 18 (extreme FID). The internal consistency of this test computed through Cronbach's alpha appeared to be acceptable (r= .75).
- The last instrument was the Raven's Advanced Progressive Matrices set II (1962). This 36-item test was used to measure the test takers' IQs. The

participants were given 50 minutes to answer the test. The number of correct responses was aggregated to achieve the test takers' IQs. This test divided the participants into four groups: Scores from 90 to 110: average; scores from 110 to 120: high-average; scores from 120 to 127: well high-average, scores above 127: superior. This test showed an acceptable estimate of reliability (r= .69).

3.3 Procedure

Senior English major students at Islamic Azad University, North Tehran Branch agreed to participate in this study. These participants were divided into two groups. One group answered the MC passage comprehension test, and the other group answered the MC rational cloze. Both groups had to accomplish their tasks in 90 minutes. The duration was expected to produce cognitive fatigue. Therefore, immediately after completing the reading test, they started to answer the CPT intended to measure their ability to sustain attention over time. Subsequent to CPT, in order to measure the participants' executive control under fatigue, once again, they answered the WCST. It is worth mentioning that prior to running the CPT and WCST, the test takers participated in a briefing session. To obtain the participants' IQs and CSs, Raven's Advanced Progressive Matrices and GEFT were administered in a different session.

4. RESULTS

To answer the first research question of the study, a one-way MANOVA was run to investigate whether there was any significant difference in the level of cognitive fatigue between the MC Passage Comprehension Group and MC Rational Cloze Group. As mentioned earlier, both groups took the WCST and CPT. Two indices were obtained through the application of WSCT: the numbers of preservative errors and completed categories. Three indices, also, were obtained from the CPT: commission errors, omission errors, and correct reaction time mean score. These five sets of scores obtained from each of the groups constituted the dependent variables. Reading comprehension testing techniques (passage comprehension and rationale cloze) were considered as the independent variable. Table 1 shows the descriptive statistics.

As Table 1 shows, 15 students took the 56-item MC passage comprehension. The same number (15) took the MC rational cloze test containing 60 items. Although both passages (comprehension and rational cloze tests) had the same readability (45.9 and 45.8, respectively), the statistical analysis in almost all cases demonstrated that rational cloze test was more difficult than the passage comprehension test. The skewness analysis (obtained by dividing the statistic of skewness by the standard error) showed that the assumption of normality was observed in the distribution of the scores (-0.03 for the Passage Group,1.25 for the Rational Cloze Group, both indices falling within the range of -1.96 and +1.96).

However, prior to running the one-way MANOVA, the preliminary assumptions for this test including normality, multivariate outliers, and homogeneity of variance-covariance was checked. In order to test the multivariate normality, Shapiro-Wilk test was used to check the normality of residuals for cognitive fatigue test scores (dependent variable) across the two reading comprehension groups (independent variable). As Table 2 shows, the scores on commission errors $p>.05$ and the correct reaction time mean $p>.05$ met the assumption of normality.

Furthermore, to investigate whether there was any outlier among the scores, the Mahalanobis distance test was used. The results revealed that there were no multivariate outlier scores across the five dependent variables (df = 5), 15.466 < 20.52. With regard to the homogeneity of variance and covariance of cognitive fatigue test scores obtained from the Passage Comprehension and Rational Cloze Groups, Box's test of equality of covariance matrices and the Levene's test of equality of error variances were run. The results are provided in Tables 3 and 4, respectively.

The results of Box's test signified that the covariance of the Passage Comprehension and Rational Cloze Groups are equal, $p>.05$. The results of the Levene's test, also, showed that the two groups had the same variances for each of the dependent variables (cognitive fatigue test scores), $p>.05$.

The results of the multivariate test, presented in Table 5, demonstrated that there was no statistically significant difference between the Passage Comprehension and Rational Cloze

Groups in terms of their cognitive fatigue, F (5, 24) = .17, p > .05; Wilks' Lambda = .97, partial eta squared = .03. Therefore, the researchers failed to reject the null hypothesis which stated that *"There is no significant difference in test takers' cognitive fatigue after their performance on passage comprehension and rational cloze".*

Subsequently, the mean scores of the dependent variables (preservative errors, completed categories, commission errors, omission errors, and correct reaction time mean) between the two groups were analyzed. The findings, presented in Table 6, confirmed that Rational Cloze Group made more preservative errors (M = 19, SD = 8.24) than the Passage Comprehension Group (M = 17.46, SD = 6.37). However, the difference was not statistically significant, F (1, 28) = .354, p > .05; partial eta squared = .01 (see Table 7). Although the increase in the number of preservative errors assumed to coincide with the increase in cognitive fatigue (Van der Linden et al., [18]), the researchers could not justify the claim because the difference in the number of the preservative errors did not reach a statistically significant level.

Table 1. Descriptive statistics for the passage comprehension and rational cloze tests

	n	K	Range	Min	Max	Mean	SD	Skewness		Cronbach-α
								statistics	std. error	
PC	15	56	36	17	53	35.40	8.48	-0.018	0.58	.86
RC	15	60	26	5	31	16.40	7.60	0.725	0.58	.77

Note. PC = passage comprehension; RC = rational cloze

Table 2. Shapiro-wilk test results of normality for residuals

Cognitive fatigue tests	Statistics	df	Sig.
Preservative Errors	.85	30	.001
Completed categories	.87	30	.002
Commission errors	.94	30	.092
Omission errors	.84	30	.000
Correct reaction time mean	.96	30	.487

p< .05

Table 3. Box's test of equality of variance-covariance matrices of dependent Variables across the groups

F	df1	df2	Sig.
1.295	15	3.15	.196

p< .05

Table 4. Test of homogeneity of variances, cognitive fatigue tests across the passage comprehension and rational cloze tests

Cognitive fatigue tests	Levene statistics	df1	df2	Sig.
Preservative Errors	.584	1	28	.451
Completed Categories	2.126	1	28	.156
Commission Errors	2.845	1	28	.103
Omission Errors	2.783	1	28	.106
Correct Reaction Time Mean	1.443	1	28	.240

p<.05

Table 5. Multivariate tests results

	Value	F	Hypothesis df	Error df	Sig.	Partial eta squared
Wilks' lambda	.96	.17	5	24	.97	.03

p < .05

Table 6. Descriptive statistics, cognitive fatigue tests across the passage comprehension & rational cloze tests

Reading tests		PE		CC		CE		OE		CRTM	
	N	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Passage	15	17.46	6.37	7.66	1.34	14.53	10.14	4.26	5.07	3.81	54.98
Cloze	15	19.06	8.24	7.20	2.11	12.93	5.49	4.06	4.06	3.88	46.60

Note. PE = preservative error, CC = completed categories, CE = commission errors, OE = omission errors, CRTM = correct reaction time mean

Table 7. Univariate test results for reading tests (passage comprehension & rational cloze tests)

Cognitive fatigue (DV)	Sum of squares	df	Mean square	F	Sig.
Preservative error	19.20	1	19.20	.35	.55
Completed categories	1.63	1	1.63	.52	.47
Commission errors	19.20	1	19.20	.28	.59
Omission errors	.30	1	.30	.019	.893
Correct reaction time mean	412.55	1	412.55	.15	.69

p<.05

Another index of cognitive fatigue which signified a small mean difference between the Rational Cloze and Passage Comprehension Groups is the number of commission errors made by the participants in the groups. An inspection of the mean scores demonstrated that Passage Comprehension Group made more commission errors (M = 14.53, SD = 10.14) than the Rational Cloze Group (M = 12.93, SD = 5.49). Although the larger number of commission errors (responses to stimuli other than the target) is frequently reported as a measure of impulsivity (Riccio, Reynolds, Lowe, & Moore, [42]), inaccuracy, and poor inhibitory control (Armengol, [43]), the finding of the present study made such interpretation impossible since the difference was not statistically significant, F (1, 28) = .289, p > .05; partial eta squared = .01 (see Table 7).

In order to answer the second research question of the study, a two-way MANOVA was run. The dependent variables were the same as in the one-way MANOVA. However, IQ, CS, and reading comprehension testing techniques were used as the independent variables. Table 8 signifies the descriptive statistics.

As shown in Table 8, 30 participants answered the 36-item IQ test; their scores ranged from 90 to 110, and their mean score (109.13) was equal to the average level. However, the mean score is close to the border line of the average level and the next level (high-average, ranging from 110 to 120). The result of skewness showed that the distribution of scores on IQ test (0.78) was within the range of ±1.96 and thus met the assumption of normality. The test also showed an acceptable level of reliability estimate (r= .69).

The same number of participants (n = 30) answered the 18-item cognitive style (GEFT) test. This test divided the test takers into two general categories of field dependent and field independent. An inspection of the mean score (12.06) showed that the test takers were more field-independent. As illustrated in Table 8, the assumption of normality was not observed, and the distribution of the scores showed high degree of negative skewness (-3.97). This finding allowed the researchers to conclude that the IQ test was to some extent difficult for the test takers, while the cognitive style test was much easier. The internal consistency on this test (r= .75), also, was acceptable. Table 9 shows the results of normality for cognitive fatigue tests residuals across the three independent variables of testing techniques, IQ, and cognitive style which are the preliminary assumptions for running MANOVA.

As Table 9 indicates, scores on omission errors p>.05 and correct reaction time mean p>.05 met the assumptions of normality. Additionally, the results of equality of variance-covariance of dependent variables (cognitive fatigue tests score) across the three independent variables (reading measurement testing techniques, IQ, and cognitive style) are presented in Table 10.

The results of Box's test suggested that the covariance of the dependent variables were the same across the three independent variables *p*>.05. The results of the Levene's test (Table 11), also, showed that all dependent variables (cognitive fatigue test scores) had the same variances across the independent variables *p*>.05.

The result of the two-way MANOVA run to examine the impact of reading measurement methods, IQ, and cognitive style on cognitive fatigue is presented in Table 12. As shown, there were no statistically significant differences among the effects of reading comprehension testing techniques F (5, 20) = .15, p >.05; Wilks' Lambda = .96; IQ F (15, 55.61) = .88, p >.05; Wilks' Lambda = .55; and cognitive style F (5, 20) = .67, p >.05; Wilks' Lambda = .85 on the test takers' cognitive fatigue. Therefore, the researchers failed to reject the second null hypothesis which stated that "*There are no significant differences among the impacts of reading measurement techniques, IQ, and cognitive style on the test takers' cognitive fatigue*".

The results of univariate test analysis for IQ and cognitive style test provided in Tables13 and 14 implied that neither the test takers' IQ, nor their cognitive styles affected cognitive fatigue.

5. DISCUSSION

The purpose of the present study was to examine whether different techniques for measuring reading comprehension ability (passage comprehension and rational cloze tests) had an effect on test takers' cognitive fatigue. More precisely, the researchers assumed that cognitive fatigue could be considered as a measurement error which interferes with measuring the construct under scrutiny and might be created due to different factors such as individual, environmental, or instrumental factors. Hence, IQ and cognitive style were investigated as individual factors and testing techniques as the instrumental factor.

As the results showed, different reading comprehension testing techniques (passage comprehension and rational cloze tests) had no significant effect on the test takers' cognitive fatigue. This finding is in line with Mayes, Valerie, Sims, and Koonce [46] who found no significant difference in test takers' workload between those who read the text from video display terminals and those who read from paper. They stated that arriving at such results might have several reasons. Their first justification was that although performance differences exist with regards to varieties of test format, people are not generally capable of perceiving these differences and in fact, they are unable to quantify the mental workload caused by different test formats. Another justification was related to the instrument (NASA Task Load Index) they used to measure test takers' mental workload. They stated that the instruments and other subjective measures of mental workload are unable to adequately distinguish the presence of an increased workload.

Table 8. Descriptive statistics for the IQ and cognitive style tests

	N	K	Range	Min	Max	Mean	SD	Skewness		Cronbach-α
								statistics	std error	
IQ	30	36	32	96	128	109.13	7.53	0.33	0.42	.69
CS	30	18	14	2	16	12.06	3.073	-1.67	0.42	.75

Note. CS = cognitive style

Table 9. Shapiro-wilk test results of normality for residuals

Cognitive fatigue tests	Statistics	df	Sig.
Preservative Errors	.88	30	.003
Completed categories	.91	30	.023
Commission errors	.89	30	.006
Omission errors	.94	30	.155
Correct reaction time mean	.96	30	.40

p<.05

Table 10. Box's test of equality of variance-covariance matrices of dependent variables

F	df1	df2	sig.
1.22	15	647.32	.249

p<.05

Table 11. Test of homogeneity of variances, cognitive fatigue tests scores across the three independent variables

Cognitive fatigue tests	Levenestatistics	df1	df2	Sig.
Preservative Errors	1.34	6	23	.280
Completed Categories	1.37	6	23	.268
Commission Errors	1.58	6	23	.198
Omission Errors	8.98	6	23	.000
Correct Reaction Time Mean	.864	6	23	.536

p<.05

Table 12. Multivariate tests

Wilks' lambda	Value	F	Hypothesis df	Error df	Sig.
Reading Tests	.96	.15	5	20	.97
IQ	.55	.88	15	55.61	.58
Cognitive Style	.85	.67	5	20	.65

p< .05

Table 13. Univariate test results for IQ test

Cognitive fatigue (DV)	Sum of squares	df	Mean square	F	Sig.
Preservative Error	38.56	3	12.84	.20	.88
Completed Categories	4.51	3	1.50	.43	.73
Commission Errors	321.44	3	107.14	1.68	.19
Omission Errors	96.48	3	32.16	2.23	.11
Correct Reaction Time Mean	7478. 75	3	2492.85	1.01	.40

p<.05

Table 14. Univariate test results for cognitive style test

Cognitive fatigue (DV)	Sum of squares	df	Mean square	F	Sig.
Preservative Error	6.40	1	6.40	.10	.75
Completed Categories	.04	1	.04	.01	.91
Commission Errors	33.90	1	33.90	.53	.47
Omission Errors	8.73	1	8.73	.60	.44
Correct Reaction Time Mean	7731.59	1	7731.59	3.14	.08

p<.05

On the other hand, the finding of this study is not consistent with the study by Van der Linden, et al. [18]. Their study showed that there were significant differences in the number of preservative errors between fatigued and non-fatigued participants on (WCST) and their planning time on Tower of London. Consistently, Barwick, Arnett, and Slobounov [47] discovered that as test takers' cognitive fatigue increased, an increasing number of errors occurred in their performances on Stroop Test and as a consequence, their accuracy decreased.

However, although no significant difference was found in the level of cognitive fatigue between the two groups of the present study, the researchers examined some of the underlying reasons which might probably cause such a result. The first reason could be because of the structure of the reading comprehension tests. As mentioned previously, the texts used for passage comprehension and rational cloze tests had the same readability (45.9 and 45.8 respectively), and the participants were at the same level of language proficiency. However, the results of the descriptive statistics (see Table1) revealed that rational cloze test (M = 16; skewness = 1.25) was more difficult than the passage comprehension test (M = 35; skewness = - 0.03). Almost all participants who answered the rational cloze tests continuously complained about the difficulty level of the test. Therefore, although the researchers' expectation was that answering the rationale cloze test might need less cognitive processing than the passage comprehension test, the difficulty of the rational cloze test might have caused an increase in the level of cognitive fatigue.

Other reasons which could have led to the insignificant difference between reading comprehension testing techniques and cognitive fatigue could presumably be due to two major limitations. The first limitation could be the small number of participants (n = 15) in each group. The second drawback of the study lay in the limited length of the fatigue-inducing process. As participants were totally reluctant to go through the fatigue manipulation process, the researchers were forced to shorten the process to 90 minutes. However, most studies reported to use more than two hours to induce fatigue in participants (Ackerman, Kanfer, Shapiro, Newton, Beier, [48]; Newton, [49]; Mizuno, Tanaka, Yamaguti, Kajimoto, Kuratsune, & Watanabe [50]; Van der Linden, et al., [18]). Thus, the short length of fatigue-inducing process could have caused testing techniques to create no difference in test takers' cognitive fatigue. Another probable reason is assumed to be the identical type of response format (multiple-choice) for both passage comprehension and rational cloze tests which might neutralize any differences of reading comprehension testing techniques.

This study further analyzed whether test takers' individual differences (here, IQ and cognitive style) contributed to the same level of cognitive fatigue. As the results revealed (see Table 10), there were no differences among the effect of reading comprehension testing techniques, IQ, and cognitive style on the test takers' cognitive fatigue. In fact, individual differences might lead to different performances on various types of tests. For example, with regard to gender differences, female students appeared to perform better on the same reading comprehension test than male students (Chiu, Chow, & McBride-Chang, [32]). Students with higher motivation to read have a higher tendency to become skilled readers compared to less motivated students (Morgan & Fuchs, [8]). The capacity of working memory according to (Daneman & Carpenter, [4]) is also one of the individual factors which affect text comprehension. Furthermore, considering the effect of IQ on test performance, as mentioned earlier, Pishghadam and Tabataba'ian [17] found no relationship between IQ and performance on MC test (except for Arithmetic sub section), while their study showed that performance on cloze test, C-test, and summary writing are related to IQ. Therefore, the present study showed that although individual differences affect test takers' performances on different types of tests, IQ and cognitive style do not appear to affect test takers' cognitive fatigue (see Tables 11 and 12).

6. CONCLUSION

The results showed that the two reading comprehension testing techniques did not differently affect test takers' cognitive fatigue. Additionally, individual differences (here, IQ and cognitive style) were not important factors in producing cognitive fatigue. However, due to the limitations in the process of the study, arriving at any interpretation or generalization should be made cautiously. Increasing the number of participants in further studies can contribute to more reliable results. Also, extending the length of fatigue-inducing process and using varieties of response formats (not merely MC formats) are other factors which can affect the results of similar studies in future. Further research can tackle other language skills and sub-skills because testing techniques might have different impacts on cognitive fatigue with regard to the language skill being measured. Finally, this study calls for further attention to how varieties of test method facets may cause different cognitive loads on test takers.

ETHICAL APPROVAL

All authors hereby declare that ethical considerations were strictly followed in conducting this study and the treatment of the participants was in accordance with the established ethical standards. It should also be noted that the participants volunteered for the study and were quite aware of the conditions of the research project. Moreover, all participants received one mark bonus for their participation in this project. Each individual was informed of the test results per their request.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Bachman LF. Fundamental considerations in language testing. Oxford: Oxford University Press; 1990.
2. Ediger A. Teaching children literacy skills in a second language. In: Celce-Murcia M, editor. Teaching English as a second or foreign language. 3rd ed. Boston: Heinle&Heinle; 2001.
3. Carretti B, Borella E, Cornoldi C, Beni RD. Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. Learning and Individual Differences. 2009;19:246–251.
4. Daneman M, Carpenter PA. Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior.1980;19(4):450–466.
5. Mih V, Mih C. The role of working memory deficits in children with poor comprehension ability. Procedia - Social and Behavioral Sciences. 2011;29:347-355.
6. Oakhill J, Hartt J, Samols D. Levels of comprehension monitoring and working memory in good and poor comprehenders. Reading and Writing. 2005;18:657–686.
7. Sellers VD. Anxiety and reading comprehension in Spanish as a foreign language. Foreign Language Annals.2000;33(5):512-520.
8. Morgan PL, Fuchs D. Is there a bidirectional relationship between children's reading skills and reading motivations? Exceptional Children. 2007;73(2):242-255.
9. Alderson JC. Assessing reading. Cambridge: Cambridge University Press; 2000.
10. Hassani L, Maasum TNRTM. A study of students' reading performance in two test formats of summary writing and open-ended questions. Social and Behavioral Sciences. 2012;69:915–923.
11. Keenan JM, Meenan CE. Test differences in diagnosing reading comprehension deficits. Journal of Learning Disabilities. 2012;47(2):125–135.
12. Kobayashi M. Method effects on reading comprehension test performance: Text organization and response format. Language Testing. 2002;19(2):193-220.
13. Han M. EFL readers' test-taking processes for completion vs. multiple choice cloze tests. The Linguistic Association of Korea Journal. 2007;15(3):189-208.
14. Kendeou P, Papadopoulos TC, Spanoudis G. Processing demands of reading comprehension tests in young readers. Learning and Instruction. 2012;22:354-367.
15. Rupp AA, Ferne T, Choi H. How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. Language Testing. 2006;23(4):441–474.
16. Ozuru Y, Dempsey K, McNamara DS. Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. Learning and Instruction. 2009;19:228-242.
17. Pishghadam R, Tabataba'ian MS. IQ and test format: A study into test fairness. Iranian Journal of Language Testing. 2011;1(1):17-29.
18. Van der Linden D, Frese M, Meijman TF. Mental fatigue and the control of cognitive processes: Effects on perseveration and planning. ActaPsychologica. 2003;113:45–65.
19. Alderson JC, Bachman LF. Assessing reading. 5th ed. Cambridge: Cambridge University Press; 2005.
20. Jafarpur AJ. Is the test constructor a facet? Language Testing. 2003;20(1):57–87.
21. Abraham RG, Chapelle CA. The meaning of cloze test scores: An item difficulty perspective. Modern Language Journal. 1992;76(4):468-479.
22. Brown JD. Do cloze tests work? Or, is it just an illusion? Second Language Studies Journal. 2002;21(1):79-125.
23. Chapelle CA, Abraham RG. Cloze method: What difference does it make? Language Testing. 1990;7(2):121-146.
24. Greene BB. Testing reading comprehension of theoretical discourse with cloze. Journal of Research in Reading. 2001;24(1):82-98.

25. Gellert AS, Elbro C. Cloze tests may be quick, but are they dirty? Development and preliminary validation of a cloze test of reading comprehension. Journal of Psychoeducational Assessment. 2013;31 (1):16–28.
26. McKameyT. Getting closure on cloze: A validation study of the “rational deletion” method. Second Language Studies. 2006;24(2):114-164.
27. Storey P. Examining the test-taking process: A cognitive perspective on the discourse cloze test. Language Testing. 1997;14(2):214-231.
28. Tajeddin Z. The relationship between the use of language learning strategies and performance on the cloze.IJAL. 2004;7(1):63-79.
29. Yamashita J. Processes of taking a gap-filling test: comparison of skilled and less skilled EFL readers. Language Testing. 2003;20(3):267–293.
30. Bensoussan M, Kreindler I. Improving advanced reading comprehension in a foreign language: Summaries vs. short-answer questions. Journal of Research in Reading. 1990;13:55–68.
31. McNamara T, Roever C. Language testing: The social dimension. Oxford: Blackwell; 2006.
32. Chiu MM, Chow B, McBride-Chang C. Universals and specifics in learning strategies: Explaining adolescent mathematics, science, and reading achievement across 34 countries. Learning and Individual Differences. 2007;17(4): 344–365.
33. Bachman LF, Palmer AS. Language testing in practice. Oxford: Oxford University Press; 1996.
34. Baker D. Language testing: A critical survey and practical guide. London: Hodder& Stoughton; 1989.
35. Grandjean EP. Fatigue. Yant Memorial Lecture. American Industrial Hygiene Association Journal. 1970;31:401-411.
36. Mizuno K, Watanabe Y. Utility of an advanced trail making test as a neuropsychological tool for an objective evaluation of work efficiency during mental fatigue. In: Watanabe Y, Evengård B, Natelson B, Jason LA, Kuratsune H, editors. Fatigue science for human health. NY: Springer; 2008.
37. Moller M. Fatigue and cognition: Hormonal perspectives. Unpublished doctoral dissertation. Karolinska Institutet: Department of Clinical Sciences, Stockholm, Sweden; 2013.
38. Nyhus E, Barceló F. The Wisconsin card sorting test and the cognitive assessment of prefrontal executive functions: A critical update. Brain and Cognition. 2009;71:437–451.
39. Somsen RJM, van der Molen MW, Jennings JR, van Beek B. Wisconsin card sorting in adolescents: Analysis of performance, response time, and heart rate. Acta Psychologica. 2000;104:27–257.
40. Barceló F, Knight RT. Both random and perseverative errors underlie WCST deficits in prefrontal patients. Neuropsychologia. 2002;40:349–356.
41. Greve KW. Can perseverative responses on the Wisconsin Card Sorting Test be scored accurately? Archives of Clinical Neuropsychology. 1993;8:511–517.
42. Riccio CA, Reynolds CR, Lowe P, Moore JJ. The continuous performance test: A window on the neural substrates for attention? Archives of Clinical Neuropsychology. 2002;17:235-272.
43. Armengol CG. Effect of co-morbid depression on continuous performance test (CPT) tasks in college students with attention deficit hyperactivity disorder. Revista Espanola de neuropsicologia. 2003;5(1):33-48.
44. Kieling C, Roman T, Doyle AE, Hutz MH, Rohde LA. Association between DRD4 Gene and Performance of Children with ADHD in a Test of Sustained Attention. Biol Psychiatry. 2006;60:1163–1165.
45. Conner CK, MHS Staff. Conners' continuous performance Test II (CPT II V.5), Profile Report. Toronto: Multi-Health Systems; 2004.
46. Mayes DK, Sims VK, Koonce JM. Comprehension and workload differences for VDT and paper-based reading. International Journal of Industrial Ergonomics. 2001;28:367–378.
47. Barwick F, Arnet P, Slobounov S. EEG correlates of fatigue during administration of a neuropsychological test battery. Clinical Neurophysiology. 2012;123:278–284.
48. Ackerman PL, Kanfer R, Shapiro SW, Newton S, Beier M.E. Cognitive fatigue during testing: An examination of trait, time-on-task, and strategy influences. Human Performance. 2010;23(5):381-402.

49. Newton SH. The effects of caffeine on cognitive fatigue. Unpublished master's thesis. Georgia Institute of Technology, US, Georgia; 2009.
50. Mizuno K, Tanaka M, Yamaguti K, Kajimoto O, Kuratsune H, & Watanabe Y. Mental fatigue caused by prolonged cognitive load associated with sympathetic hyperactivity. Behavioral and Brain Functions. 2011;7:17.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history.php?iid=820&id=21&aid=7756