



## Answer Changing in Online and Traditional Pen-Paper Tests: The Case of Upper Intermediate EFL Learners with Different Cognitive Styles

Fatemeh Zolfaghari \* 

*Assistant Professor, Center of English Language, Isfahan University of Technology, Isfahan, Iran*

Received: 2023/07/09

Accepted: 2023/08/21

**Abstract:** The present study addressed the effects of cognitive styles on answer-switching practices across online and traditional tests. After completing Ehrman and Leaver's (2003) cognitive style questionnaire, a sample of upper intermediate students took pen-paper and online versions (59 test-takers each) of an already validated teacher-developed test of English. The data from think-aloud and erasure analyses revealed significantly more frequent total and right-to-wrong changes in the traditional and online tests, respectively. Multiple regression values explained more than 50% of right/wrong-to-wrong, wrong-to-right, and overall answer-changing variance based on cognitive styles in the pen-paper exam. However, the regression results from online performance analyses could not prove the power of thinking styles in predicting answer-changing strategies. Fisher's exact tests showed significantly different answer-changing strategies adopted by field-dependent, leveler, analog, concrete, and impulsive individuals in the traditional test but no significant differences between the behaviors of individuals with different cognitive styles in the online exam. Based on the present findings, online and pen-paper platforms may require different test-taking strategies. Language instructors and test developers can use these findings to align their instructional and assessment practices with various cognitive styles and testing environments.

**Keywords:** Answer Changing, Online Tests, Pen Paper Tests, EFL Learners, Cognitive Styles.

---

\* Corresponding Author.

Authors' Email Address:

Fatemeh Zolfaghari (fzolfaghari@iut.ac.ir)

2322-5343 © University of Isfahan

This is an open access article under the CC BY-NC-ND/4.0/ License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).



## Background

Standardized multiple-choice (MC) tests have the potential to provide sufficient stimuli to unleash human capabilities and reduce unfair educational decisions (Armstrong, 1993; Ennis, 1993; Haataja et al., 2023; Lau et al., 2011; Phelps, 2009; Zaidi et al., 2018). Hence, despite a lack of consensus about the most efficient strategies to answer discrete-point items and achieve valid results from knowledge assessment, MC tests still serve as efficient means in large-scale language testing (Haataja et al., 2023; Sellah et al., 2018). However, it is still uncertain whether test-taking strategies are equally effective as the depth and breadth of content knowledge in achieving success on traditional and online language tests (Aryadoust, 2019; Collier, Pillai, & Fazio, 2023).

MC tests may create the conditions for using answer-changing strategies and engaging in cognitive activities. In response to multiple-choice items, test takers may show various dispositions. They may have full knowledge of the assessed content and mark the correct option almost certainly. When there is a knowledge deficit, they may omit the options, merge their disparate estimates, or follow their hunches (Herzog & Hertwig, 2014). The latter test-taking strategies have prompted studies on answer-changing behaviors.

Previous studies have yielded conflicting findings on whether answer-switching behaviors contribute to score gains on MC tests. A significant number of studies have given credit to the contribution of double thoughts to better test scores in traditional pen-paper (e.g., Merry et al., 2021; Stylianos-Georgiou & Papanastasiou, 2017) and computer-based (e.g., Liu et al., 2015; Vispoel, 1998) tests. However, considering the dynamic and context-dependent nature of required strategic choices in achieving and showcasing language skills, it seems unrealistic to introduce a uniform policy for everyone (Chen et al., 2014; Morrison, 1988). Hence, language testing and assessment scholars have started focusing on how test-takers' increased awareness of their unique cognitive abilities (e.g., cognitive styles) impacts their performance across various assessment platforms (Couchman et al., 2016; Herzog & Hertwig, 2014; Stylianos-Georgiou & Papanastasiou, 2017).

As a reflection of Gestalt and Piagetian theories of cognitive development, cognitive styles embody a set of mental operations such as decision-making, reasoning, judgment, and problem-solving (Pitcher, 2002; Riding & Rayner, 1998), each regulating reception of stimuli and transformation of knowledge into quick solutions to challenging tasks (Zhang, 2023). Awareness

of the potential workings of cognitive styles on test-taking performance can direct stakeholders to make appropriate instructional decisions, pursue suitable assessment policies (Griffiths, 2012), and predict the chances for success in cognitive and learning tasks (Parry, 1984). Also, strong familiarity with various mental operations and preferences of language test takers can maintain harmony between instructional and assessment practices and maximize their effectiveness (Cohen & Weaver, 2006; Griffiths, 2012; Zhang, 2023). By implication, test takers and stakeholders can align their time and efforts with the difficulty levels of tasks (Efklides, 2012).

Accordingly, a dearth of studies on answer changing in online testing platforms, conflicting views on the possible effects of answer-switching, and individual differences in score gains/losses in pen-and-paper tests call for further investigations into strategies for taking MC tests (Dodeen, 2008; Peng et al., 2014; Geiger, 1991; Papanastasiou & Reckase, 2008). To address some of these gaps, this study investigates the predictive power of cognitive styles in determining EFL learners' answer changes on online and traditional forms of teacher-developed English achievement tests. In so doing, the following research questions are examined:

1. Do answer-changing patterns significantly differ across test takers with different cognitive styles?
2. Do answer-changing patterns significantly differ across online and traditional pen-and-paper tests?
3. Do cognitive styles predict answer changing in English achievement tests across online and traditional pen-and-paper platforms?

## **Method**

### ***Participants***

Based on availability sampling to recruit readily accessible volunteers, this study employed a convenient sample of 50 male and 68 female university students with upper intermediate levels of general English proficiency and an age range of 19 to 25 who pursued diverse engineering disciplines. The participants had not pursued any additional English education and qualifications outside the English courses offered in the mainstream school system before university. They reported mild or almost no anxiety before and during the test and had spent 2.5 weeks on average preparing for the test.

### *Instruments*

The quick Oxford Placement Test (OPT) was the first instrument to check and control for the participants' proficiency levels. Already identified as a valid test (Geranpayeh, 2003) and with a high Kuder-Richardson 21 reliability index of .87 in the present study, OPT has proved successful in controlling the effects of proficiency levels on other variables (e.g., Abdulhussien, 2023; Ashraf Nia et al., 2023; Azizi & Nemati, 2022). The first 40 OPT items are appropriate for learners with different proficiency levels, but the last 20 questions suit learners with upper intermediate and advanced proficiency levels. The study sample comprised participants with OPT scores between 40 and 47 (i.e., B2 or upper intermediate level).

Secondly, a validated teacher-developed English achievement test was applied to detect the tendencies to change answers on MC items. The initial version of the test consisted of 60 items (i.e., 15 grammar and 45 reading comprehension items) based on the language contents taught during a 4-month academic semester. Four university professors with 5 to 20 years of experience in Teaching English as a Foreign Language commented on the content and face validity of the 60-item test. Several rounds of factor analyses using SPSS 22 helped satisfy the construct validity of the items. Accordingly, with 15 items left out, the validated test consisted of 45 items loading on the underlying components (i.e., 30 reading comprehension and 15 grammar items). A satisfactory index of .84 also helped ensure the internal consistency of the 45-item test (Pallant, 2005).

The third instrument was the Persian version of the cognitive style questionnaire (Ehrman & Leaver, 2003) validated by Maftoon and Rezaie (2013) for the Iranian context. The questionnaire consisted of 30 nine-point items targeting field-dependent/field-independent (items 1, 11, and 21), sharpener/leveler (items 3, 13, and 23), impulsive/reflective (items 5, 15, and 25), global/particular (items 4, 14, and 24), analog/digital (items 7, 17, and 27), concrete/abstract (items 8, 18, and 28), field-sensitive/field-insensitive (2, 12, and 22), synthetic/analytic (items 6, 16, and 26), random/sequential (items 9, 19, and 29), and inductive/deductive (items 10, 20, and 30) styles.

### *Procedures*

Administration of Oxford Quick Placement Test, which took about 30 minutes, helped identify 118 EFL learners with upper-intermediate levels of English proficiency as the study participants. The participants also needed to take the teacher-developed English achievement test on grammar and reading comprehension. The next phase of the study involved checking fully filled-out

answer sheets with no missing items coupled with the recorded think-aloud videos of the test takers and collecting the overall and separate records of answer changes in each test section. As there were no penalties for incorrect responses and the participants could turn to their prior knowledge in cases of uncertainty, the number of missing items was low, but probably with a high likelihood of blind guessing. However, even the participants' blind guesses could not seriously damage the collected data since they could not reconsider their choices without content knowledge or mental operations (Geiger, 1991).

The collected data comprised right-to-wrong and wrong-to-right answer changes based on the observed erasure behaviors. To reduce the likelihood of mental answer changes, the course instructor provided general guidelines asking the pen-and-paper test takers to make their changes observable through erasure. Each test taker was supposed to report their changes by recording their choices with a fountain pen and crossing out the previously selected options in case of further changes. Training the students through the course sessions on marking their mental answer changes could remedy the methodological deficiencies of the previous studies due to their complete reliance on ex post facto analyses of erasure behaviors in answer sheets. Besides, checking the reliability index and performing several rounds of factor analysis to confirm item appropriateness could reduce the chances of extremely easy or difficult test items; hence, the frequencies of the students' back-and-forth work in marking options could be considered reasonable. The traditional test takers were also allowed a time limit for transferring the responses to the answer sheets. For the online version of the same test, the participants installed screen recorder software and thought out loud while taking the test. Then, they shared the recorded files for further analysis.

Before proceeding with the main study, four experts checked the questionnaire items and helped make minor modifications to their formatting. Then, to ensure the appropriateness of the scale for the study purposes, decide on the average questionnaire completion time, and identify and resolve the potential issues, a pilot analysis involving a representative sample of 33 participants from the target population completed the questionnaire. The results indicated that discarding or modifying none of the items could increase the reliability index of the scale (Dornyei, 2010), and on average, each participant needed about 12 minutes to fill out the questionnaire. The Cronbach's alpha reliability for the internal consistency of the questionnaire items equaled .81, which was highly satisfactory (Pallant, 2005).

During the data analysis phase, Fisher's exact tests compared answer-switching across various cognitive styles. Subsequently, multiple linear regression analyses gave the required

indices to check the predictive power of cognitive learning styles in determining the types and frequencies of answer-changing behaviors.

## Results

### *Answer Changing in the Online Test*

Fisher's Exact Test helped compare the frequencies of total, wrong-to-right, and right-to-wrong changes across online test takers with different cognitive styles (Table 1). The results indicated that reflective and digital test-takers tended to change their answers more frequently than their impulsive and analog counterparts.

**Table 1.** Comparison of Answer Changing across Different Cognitive Styles in the Online Test

Styles	N	Total changes		Changes to right		Changes to wrong						
		F	Fisher's		Fisher's		Fisher's Exact		F	Test		
			Exact Test	F	Exact Test	F	Value	Sig.*			Value	Sig.*
Field-dependent	35	152	8.87	.56	91	5.72	.76	61	11.01	.06		
Field-independent	24	119			58			61				
Leveler	29	122	10.47	.36	73	9.87	.21	49	8.61	.16		
Sharpener	30	149			76			73				
Impulsive	28	124	15.70	<b>.05</b>	66	7.12	.55	58	6.87	.31		
Reflective	31	147			83			64				
Global	14	67	4.43	.98	32	9.44	.24	35	5.77	.45		
Particular	45	204			117			87				
Analog	30	123	15.88	<b>.05</b>	69	7.35	.52	54	5.66	.48		
Digital	29	148			80			68				
Concrete	39	181	9.89	.43	101	5.72	.73	80	5.08	.55		
Abstract	20	90			48			42				
Field-sensitive	38	174	9.51	.48	94	7.71	.45	80	8.64	.16		
Field-insensitive	21	97			55			42				
Synthetic	13	57	10.35	.36	25	5.52	.72	32	5.25	.52		
Analytic	46	214			124			90				
Random	33	142	11.06	.30	80	6.48	.64	62	8.16	.20		
Sequential	26	129			69			60				
Inductive	33	139	12.28	.20	76	8.68	.33	63	7.95	.21		
Deductive	26	132			73			59				

\*: 2-sided; N: number of participants; F= frequencies of changes

### *The Predictive Power of Cognitive Styles on Answer Changing in the Online Test*

Multiple linear regression analyses assisted in specifying if cognitive styles could predict the answer-changing patterns in the online test. As the ANOVA table suggests, the F values for none of the changes were significant ( $\text{sig.} = .00 > .05$ ), indicating the inability of the regression model to explain test takers' answer-changing practices.

**Table 2.** *The Fitness of the Regression Model for the Online Test Data*

	<b>Models</b>	<b>Sum of squares</b>	<b>df</b>	<b>Mean square</b>	<b>F</b>	<b>Sig.</b>
<b>Regression</b>	<b>Change to wrong</b>	27.732	10	2.773	1.331	.242
	<b>Change to right</b>	16.342	10	1.634	.455	.910
	<b>Total</b>	29.113	10	2.911	.548	.847
<b>Residual</b>	<b>Change to wrong</b>	99.997	48	2.083		
	<b>Change to right</b>	172.370	48	3.591		
	<b>Total</b>	255.124	48	5.315		
<b>Total</b>	<b>Change to wrong</b>	127.729	58			
	<b>Change to right</b>	188.712	58			
	<b>Total</b>	284.237	58			

### *Answer Changing in the Traditional Test*

The Fisher's Exact Test analyses of answer-switching behaviors in the traditional pen-and-paper test suggested significantly more frequent total answer changing by field-dependent, leveler, and impulsive test takers, significantly more frequent wrong-to-right changes by field-dependent and impulsive test takers, and significantly more frequent right-to-wrong changes by impulsive, analog and concrete test takers compared with their counterparts who stood at the opposite ends of cognitive style continua (Table 3).

**Table 3.** Comparison of Answer Changing across Different Cognitive Styles in the Traditional Test

Styles	N	Total changes		Changes to right		Changes to wrong				
		F	Fisher's		F	Fisher's		F	Fisher's	
			Exact Test	Value		Sig.*	Exact Test		Value	Sig.*
Field-dependent	34	242			189			53		
Field-independent	25	78	25	.01	52	20.81	.02	26	7.83	.19
Leveler	27	169	20.80	.04	128	15.05	.23	41	4.04	.75
Sharpeners	32	151			113			38		
Impulsive	29	259	52.55	.00	190	36.26	.00	69	38.76	.00
Reflective	30	61			51			10		
Global	19	108	12.74	.53	87	11.59	.57	21	4.40	.69
Particular	40	212			154			58		
Analog	29	198	17.91	.14	143	17.59	.09	55	11.94	.03
Digital	30	122			98			24		
Concrete	30	174	15.95	.26	132	11.33	.63	42	12.77	.02
Abstract	29	146			109			37		
Field-sensitive	38	214	19.22	.07	164	17.29	.09	50	6.38	.35
Field-insensitive	21	106			77			29		
Synthetic	30	132	15.50	.31	101	9.40	.84	31	6.83	.29
Analytic	29	188			140			48		
Random	26	127	8.17	.96	90	11.29	.62	37	6.86	.30
Sequential	33	193			151			42		
Inductive	31	193	16.66	.21	147	14.45	.28	46	3.11	.91
Deductive	28	127			94			33		

\*: 2-sided; N: number of participants; F= frequencies of changes

### ***The Predictive Power of Cognitive Styles on Answer Changing in the Traditional Test***

The ANOVA table from multiple linear regression analyses suggested significant F values for wrong-to-right, right-to-wrong, and total changes (sig.= .00<.05), indicating the fitness of the regression model for the data and 95% ability of the model to explain all the answer changing practices (Table 4).

**Table 4.** *The Fitness of the Regression Model for the Traditional Test Data*

	Models	Sum of squares	df	Mean square	F	Sig.
<b>Regression</b>	<b>Change to wrong</b>	77.56	10	7.76	5.67	<b>.00</b>
	<b>Change to right</b>	391.78	10	39.18	5.94	<b>.00</b>
	<b>Total</b>	752.39	10	75.24	15.43	<b>.00</b>
<b>Residual</b>	<b>Change to wrong</b>	65.67	48	1.368		
	<b>Change to right</b>	316.80	48	6.60		
	<b>Total</b>	234.02	48	4.88		
<b>Total</b>	<b>Change to wrong</b>	143.22	58			
	<b>Change to right</b>	708.58	58			
	<b>Total</b>	986.41	58			

As the independent variables (i.e., cognitive styles) did not correlate strongly and all Tolerance, as well as VIF values, were respectively above .10 and below ten, the possibility of co-linearity was ruled out. Ultimately, as none of the independent variables could change the prediction power of the model, all of them were retained. According to the significance and Beta values (Appendix A, B, & C), impulsive/reflective styles could predict right-to-wrong (76% of the variance) and wrong-to-right (58% of the variance), as well as total changes (78% of the variance), and inductive/deductive test takers, could significantly predict total answer changes (18% of the variance) ( $\text{sig} < .05$ ). Altogether, the regression model of the study explained 54% of the right-to-wrong, 55% of wrong-to-right, and 76% of the total answer changing variance (Table 5).

**Table 5.** *Regression Model Summary for the Traditional Test*

Models	R	R <sup>2</sup>	Adjusted R <sup>2</sup>	Std. error of the estimate	Change Statistics				
					R <sup>2</sup> change	F change	Df1	Df2	Sig. F change
<b>R/W</b>	.74	.54	.45	1.17	.54	5.67	10	48	<b>.00</b>
<b>W/R</b>	.74	.55	.46	2.57	.55	5.94	10	48	<b>.00</b>
<b>Total</b>	.87	.76	.71	2.21	.76	15.43	10	48	<b>.00</b>

Predictors (constant): style 1, style 2, style 3, style 4, style 5, style 6, style 7, style 8, style 9, style 10

Dependent variable: right-to-wrong (model 1); wrong-to-right (model 2); and total (model 3) answer changes

### *Comparison of Answer Changing in Online and Traditional Tests*

Through the final phase of the analysis, the test takers' answer-changing behaviors across online and traditional pen-paper platforms were also investigated. The results indicated that traditional test takers were more likely to change their responses. However, right-to-wrong changes by online test takers were significantly more frequent.

**Table 6.** Comparison of Answer Changing across Online and Traditional Tests

Test	Total changes		Changes to right		Changes to wrong					
			Fisher's Exact		Fisher's Exact					
Formats	N	F	Test	F	Test	F	Test			
			Value	Sig.	Value	Sig.	Value	Sig.		
Online	59	271	28	.01	149	14.54	.28	122	16.23	.01
Traditional	59	320			241			79		

\*: 2-sided; N: number of participants; F= frequencies of changes

### **Discussion**

The descriptive and regression results suggested that, unlike the online platform, thinking styles could explain the participants' answer changes in the traditional pen-and-paper test. Impulsive-reflective modes of thinking could predict over half of the wrong-to-right, right-to-wrong, and total answer-changing practices in the pen-and-paper platform. With a much lower percentage, the inductive-deductive pair also proved beneficial in predicting the overall answer switches.

### *Answer Changing in the Online Test*

In the online version of the test, reflective and digital test takers outperformed their impulsive and analog counterparts in their total answer-switching practices. Also, cognitive styles could not predict online answer changes. Since individuals with a reflective cognitive style tend to be more analytic, hesitant, and accurate than their impulsive counterparts (Estaji & Safari, 2023; Rosencwajg & Corroyer, 2005), the overall changes confounded the general expectations. However, as reflective learners may perform better on analytic items and face challenges with global ones (Zelniker & Jeffrey, 1976), a possible justification can be the greater frequency of reading comprehension items covering two-thirds of the total questions and being twice as frequent as the vocabulary items.

Also, because of no significant differences between the wrong-to-right changes (i.e., score gains) made by reflective and impulsive individuals, the implication is that back-and-forth movements by reflective learners between analytic items have made up for the score loss on global ones. Another factor worth considering was the online test duration that, although similar to the traditional test, due to the possibility of unstable internet connections or inconsistent internet speed, could constrain reflective individuals who mainly need more time latency (Davoudi & Heydarnejad, 2020; Estaji & Safari, 2023).

In line with the expectations, digital individuals, who, by definition, were less reflective, tended not to make inferences about the given information, mainly focused on the surface forms, and changed their responses more frequently (Ehrman & Leaver, 2003). Similar behaviors of digital and reflective test takers with rather opposite thinking modes indicate that regardless of their dispositions to make accurate guesses about the correct item responses, test takers seem not to benefit from answer changes in online platforms.

### ***Comparison of Answer Changing in Online and Traditional Tests***

Disregarding the cognitive style differences, the significantly more frequent right-to-wrong changes in the online exam reflect the online test takers' unwise decisions, thus suggesting that answer changing cannot necessarily do the test takers any good if they take online tests. Therefore, it would be inadvisable to change answers except for the cases with misunderstood or misinterpreted item stems (Aryadoust, 2019; Ramsey, Ramsey et al., 1987). However, in the traditional test, total and wrong-to-right answer switching was significantly more frequent, pointing to the already proposed claims stressing the beneficial effects of answer changing. Aligned with this finding, Geiger (1991) argued that second thoughts on item responses also have wrong-to-wrong and right-to-wrong manifestations, but the gains from wrong-to-right changes can make up for the lost points.

### ***Answer Changing in the Traditional Pen-and-Paper Test***

In the traditional platform, field-dependent, leveler, and impulsive learners tended to make more total changes. field-dependent and impulsive learners made more wrong-to-right changes, and those with analog, impulsive, and concrete cognitive styles made more right-to-wrong changes. The field-dependent learners' frequent wrong-to-right changes supported previous notes on teacher-developed achievement tests and more or less reflected the overreliance of field-dependent test-takers on contextual cues (Richards et al., 1992).

Moreover, the findings indicating no significant differences in answer-changing behaviors of field-sensitive and field-insensitive test-takers could, unlike the general assumptions, give credit for consideration and treatment of field-insensitivity and field independence as distinct but interrelated thinking styles, thereby signifying the combined effects of field-dependence and insensitivity on this group's frequently produced changes (Ehrman & Leaver, 2003). The results, however, could not support previous theoretical arguments on the maximum performance of learners with strong field-independent and field-sensitive tendencies (e.g., Angeli, 2013; Davis & Frank, 1979; Ehrman & Leaver, 2003).

One reason for the differences between impulsive and reflective test takers in answer-changing behaviors may be the different mechanisms and reasons they have followed to revisit their responses (Zhang, 2023). Previous studies have shown that misreading the questions and reconceptualizing item stems and the requested information were two main reasons for answer-changing practices (Kruger et al., 2005; Schwartz et al., 1991), an argument quite representative of impulsive individuals. Because of their quick reading habits, impulsive test takers may have needed to reread the item stems, especially the tricky ones, to ensure they have extracted the requested information, thereby increasing the probability of changing their responses. However, reflective learners' successful changes could have been due to the positive effects of their delayed decisions on the accuracy of their choices (Koriat, 2012). A justification for the right-to-wrong changes of this second group can be the possible effects of the time they lost retrieving some further information. Accordingly, being reflective does not guarantee that a test taker always makes successful judgments because other factors, such as item difficulty parameters, can affect the results. For example, in the case of difficult questions, it is highly likely that even by reconsideration and reconceptualization of the items, test takers, regardless of their amount of care and attention, make incorrect decisions (Efklides, 2012).

In this study, wrong-to-right changes were compatible with the digital mode of thinking. Given that the consulted texts in the class and exam sessions did not resource any literary content, getting the gist of the information did not require any creative strategies. Surface strategies such as memorizing grammatical points, literal meanings, and established reading techniques could serve as assets to find the correct answers. This situation was probably the most satisfactory for digital individuals with the tendency for part-to-whole analyses of reading passages, sequential and logical approaches, and shallow surface structures (Ehrman & Leaver, 2003).

A common feature of field-dependent, impulsive, digital, and abstract individuals is that they are likely to process the content holistically (Ehrman & Leaver, 2003; Rozencaj & Corroyer, 2005). The results pointing to the more frequent incorrect changes in analog and concrete individuals, as opposed to digital and abstract ones, together with more successful changes on the part of field-dependent people, may support the conclusion that wholistic learners and those who are after the gist of issues can act more successfully in traditional answer changing attempts (Richards et al., 1992). Literature also supports that the increased frequency of changes increases the chances for compensating score loss due to wrong choices (Geiger, 1991). However, the present results do not support previous theoretical arguments (e.g., Angeli, 2013; Davis & Frank, 1979; Ehrman & Leaver, 2003) on the maximum performance of learners with strong field-independence tendencies.

Some studies on pen and paper tests suggested that greater reliance on first hunches on MC items is generally closer to the correct answers (Pressley et al., 1990). However, this study showed that the initial choices are not necessarily acceptable in all situations (Pressley & Ghatala, 1988). At odds with the online platform, the traditional test practices indicated a positive association between increased accuracy and one's attempts to monitor the responses (Efklides, 2012). Hence, a wise policy to decide whether to change their initial responses or keep to them is avoiding both hypercorrections and overconfidence (Efklides, 2012; Metcalfe & Finn, 2012). Herzog and Hertwig (2014) believed that in case of contradictory estimates, test takers should resort to dialectical bootstrapping to take advantage of averaging their rough guesses about the possible correct responses to an item (Herzog & Hertwig, 2014). Based on this conceptualization, answer changing can be beneficial because the test takers think over their choices and reconsider them to find logical and convincing reasons underlying the choice of options, and each time, they can analyze the item from a different perspective, hence a reasonable time delay, activation of the already stored knowledge, and the lower possibility of errors (Vul & Pashler, 2008).

## Conclusion

The present findings indicated that test takers benefit from reconsidering their choices upon doubt on traditional MC tests, thus implicitly showing a close relationship between test scores and the mere number of answer changes (e.g., Reiling & Taylor, 1972). However, given the poor answer-switching endeavors on the online platform, the findings can allow the test developers to consider the intervening variables (e.g., exam duration) involved in online

exams (Vul & Pashler, 2008). Also, due to the possible adoption of unethical test-taking strategies such as cheating, putting any interpretations on any sets of the obtained results requires caution (Ellenburg, 1973). Hence, further studies can, on the one hand, reduce the possibility of examination offenses due to an extended exam time and incorrect conceptualizations due to limited exam time on the other. Altogether, as long as answer changes are not affected by educated or informed guesses, they are less likely to improve exam results and thus are not recommended.

The results of this study can raise test takers' awareness of the test-taking strategies that suit their cognitive styles to perform more successfully in online and traditional tests. The results can further help test developers resolve the issues (e.g., individual traits) regulating test performance. They can utilize the present findings to ensure fair assessment of language skills and create more standardized tests. The instructors can also use the results to offer wiser advice to test takers and shift their mere focus away from task characteristics toward learner-and-task features.

Future studies can address answer changes in light of the regulatory effects of other psychological factors, such as self-awareness and self-confidence. Previous studies have shown that while revising their choices, people tend to rely more on themselves and take advice that agrees with their initial decision (Bonaccio & Dalai, 2006; Herzog & Hertwig, 2014; Soil & Larrick, 2009; Yaniv & Milyavsky, 2007). Hence, it cannot be easy to advise the students to reconsider their choices during the exam. Other issues worth investigating are whether giving general guidelines before the exam or item-specific advice during the exam can work better and if the characteristics of feedback providers (e.g., self-confidence) affect answer changing and MC test scores. Given that low achievers are highly likely to experience lower levels of self-awareness, overconfidence, and biased responses and tend not to change their options upon further reconsideration (Kruger & Dunning, 1999; Stylianou-Georgiou & Papanastasiou, 2017), they can be a potential target for future analyses. It should also be noted that gender was not included as a variable in this study. Therefore, it is recommended that other researchers take it into account in their future investigations.

### **Acknowledgment**

I would like to express my special thanks to the editorial board and anonymous reviewers of the Applied Research on English Language journal.

## References

- Abdulhussien, S. H. (2023). The impact of synchronous online teaching on Iraqi EFL learners' oral comprehension. *Applied Research on English Language*, 12(2), 1-18. <https://doi.org/10.22108/are.2022.134671.1967>
- Angeli, C. (2013). Examining the effects of field dependence-independence on learners' problem-solving performance and interaction with a computer modeling tool: Implications for the design of joint cognitive systems. *Computers & Education*, 62, 221-230. <https://doi.org/10.1016/j.compedu.2012.11.002>
- Armstrong, A. M. (1993). Cognitive Style Differences in Testing Situations. *Educational Measurement: Issues and Practice*, 12(3), 17-22. <https://doi.org/10.1111/j.1745-3992.1993.tb00538.x>
- Aryadoust, V. (2019). Dynamics of item reading and answer changing in two hearings in a computerized while-listening performance test: An eye-tracking study. *Computer Assisted Language Learning*, 33(5-6), 510-537. <https://doi.org/10.1080/09588221.2019.1574267>
- Ashraf Nia, R., Roohani, A., & Hashemian, M. (2023). Exploring implicit and explicit lexical strategies in 12 learners' incidental vocabulary learning while reading. *Applied Research on English Language*, 12(1), 133-158. <https://doi.org/10.22108/are.2022.134377.1958>
- Azizi, M., & Nemati, M. (2022). Reinforced teacher corrective feedback and learners' use of subordination clauses. *Applied Research on English Language*, 11(3), 121-142. <https://doi.org/10.22108/are.2022.131955.1825>
- Bonaccio, S., & Dalai, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, 101(2), 127-151. <https://doi.org/10.1016/j.obhdp.2006.07.001>
- Chen, Y., Lin, C., & Lin, S. (2014). EFL learners' cognitive styles as a factor in the development of metaphoric competence. *Journal of Language Teaching and Research*, 5, 698-707. <https://doi.org/10.4304/JLTR.5.3.698-707>
- Cohen, A. D. & Weaver, S. J. (2006). *Styles and strategies-based instruction: A teachers' guide*. MN: Center for Advanced Research on Language Acquisition, University of Minnesota.

- Collier, J. R., Pillai, R. M. & Fazio, L. K. (2023). Multiple-choice quizzes improve memory for misinformation debunks, but do not reduce belief in misinformation. *Cogn. Research*, 8(37). <https://doi.org/10.1186/s41235-023-00488-9>
- Couchman, J. J., Miller, N. E., Zmuda, S. J., Feather, K., & Schwartzmeyer, T. (2016). The instinct fallacy: The metacognition of answering and revising during college exams. *Metacognition & Learning*, 11, 171–185. <https://doi.org/10.1007/s11409-015-9140-8>
- Davis, J. K., & Frank, B. M. (1979). Learning and memory of field independent-dependent individuals. *Journal of Research in Personality*, 13(4), 469-479. [https://doi.org/10.1016/0092-6566\(79\)90009-6](https://doi.org/10.1016/0092-6566(79)90009-6)
- Davoudi, M., & Heydarnejad, T. (2020). The interplay between reflective thinking and language achievement: a case of Iranian EFL learners. *Language Teaching Research Quarterly*, 18, 70-82.
- Dodeen, H. (2008). Assessing test-taking strategies of university students: Developing a scale and estimating its psychometric indices. *Assessment & Evaluation in Higher Education*, 33, 409-419. <https://doi.org/10.1080/02602930701562874>
- Dornyei, Z. (2010). *Questionnaires in second language research: Construction, administration, and processing*. Routledge.
- Ehrman, M. E., & Leaver, B. L. (2003). Cognitive styles in the service of language learning. *System*, 31, 391-415. [https://doi.org/10.1016/S0346-251X\(03\)00050-2](https://doi.org/10.1016/S0346-251X(03)00050-2)
- Efklides, A. (2012). Commentary: How readily can findings from basic cognitive psychology research be applied in the classroom? *Learning and Instruction*, 22, 290-295. <https://doi.org/10.1016/j.learninstruc.2012.01.001>
- Ellenburg, F. C. (1973). Cheating on tests: Are high achievers greater offenders than low achievers? *Clearing House*, 47 (7), 427-29.
- Ennis, R. H. (1993). Critical thinking assessment. *Theory into Practice*, 32(3), 179-186. <https://doi.org/10.1080/00405849309543594>
- Estaji, M., & Safari, F. (2023). Learning-oriented assessment and its effects on the perceptions and argumentative writing performance of impulsive vs. reflective learners. *Language Testing in Asia*, 13(31). <https://doi.org/10.1186/s40468-023-00248-y>
- Geiger, M. A. (1991). Changing multiple-choice answers: Do students accurately perceive their performance? *The Journal of Experimental Education*, 59(3), 250-257. <https://doi.org/10.1080/00220973.1991.10806564>

- Geranpayeh, A. (2003). A quick review of the English quick placement test. *Research Notes Quarterly*, 12(3), 8-10.
- Griffiths, C. (2012). Learning styles: Traversing the quagmire. In S., Mercer, S., Ryan, & M. Williams (Eds.), *Psychology for language learning: Insights from research, theory and practice* (pp., 151-168). Palgrave Macmillan.
- Haataja, E. S. H., Tolvanen, A., Vilppu, H., Kallio, M., Peltonen, J., & Metsapelto, R. (2023). Measuring higher-order cognitive skills with multiple choice questions –potentials and pitfalls of Finnish teacher education entrance. *Teaching and Teacher Education*, 122, 103943. <https://doi.org/10.1016/j.tate.2022.103943>
- Herzog, S. M., & Hertwig, R. (2014). Think twice and then: Combining or choosing in dialectical bootstrapping? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 218–232.
- Koriat, A. (2012). The relationships between monitoring, regulation and performance. *Learning and Instruction*, 22, 296–298. <https://doi.org/10.1016/j.learninstruc.2012.01.002>
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77, 1121–1134. <https://doi.org/10.1037//0022-3514.77.6.1121>
- Kruger, J., Wirtz, D., & Miller, D. T. (2005). Counterfactual thinking and the first instinct fallacy. *Journal of Personality and Social Psychology*, 88, 725–735. <http://dx.doi.org/10.1037/0022-3514.88.5.725>
- Lau, P. N. K., Lau, S. H., Hong, K. S., & Usop, H. (2011). Guessing, partial knowledge, and misconceptions in multiple-choice. *Journal of Educational Technology & Society*, 14 (4), 99–110.
- Liu, O. L., Bridgeman, B., Gu, L., Xu, J., & Kong, N. (2015). Investigation of response changes in the GRE Revised General Test. *Educational and Psychological Measurement*, 75, 1002–1020. <https://doi.org/10.1177/0013164415573988>
- Maftoon, P., & Rezaie, G. (2013). Cognitive style, awareness, and learners' intake and production of grammatical structures. *Journal of Language and Translation*, 3(3), 1-15.
- Merry, J. W., Elenchin, M. K., & Surma, R. N. (2021). Should students change their answers on multiple choice questions? *Adv Physiol Educ*, 45, 182-190. <https://doi.org/10.1152/advan.00090.2020>

- Metcalf, J., & Finn, B. (2012). Hypercorrection of high confidence errors in children. *Learning and Instruction*, 22(4), 253-261. <https://doi.org/10.1016/j.learninstruc.2011.10.004>
- Morrison, D. L. (1988). *Predicting diagnosis performance with measures of cognitive style. Current Psychology*, 7(2), 136-156. <https://doi.org/10.1007/BF02686657>
- Pallant, J. (2005). *SPSS survival manual: A step by step guide to data analysis using SPSS for Windows*. Allen & Unwin.
- Papanastasiou, E. C., & Reckase, M. D. (2008). *Item review as a non-traditional method of item analysis*. Annual Conference of the Psychometric Society, Durham, NH.
- Parry, T. S. (1984). *The relationship of selective dimensions of learner cognitive style, aptitude, and general intelligence factors to selected foreign language proficiency tasks of second-year students of Spanish at the secondary level* [Doctoral Dissertation, The Ohio State University].
- Peng, Y., Hong, E., & Mason, E. (2014). Motivational and cognitive test-taking strategies and their influence on test performance in mathematics. *Educational Research and Evaluation*, 20(5), 366–385. <http://dx.doi.org/10.1080/13803611.2014.966115>
- Phelps, G. (2009). Just knowing how to read isn't enough! Assessing knowledge for teaching reading. *Educational Assessment, Evaluation and Accountability*, 21, 137-154. <https://doi.org/10.1007/s11092-009-9070-6>
- Pitcher, R. T (2002). Cognitive learning styles: A review of the field dependent- field independent approach. *Journal of Vocational Education and Training*, 54 (1), 177-132. <https://doi.org/10.1080/13636820200200191>
- Pressley, M., & Ghatala, E. S. (1988). Delusions about performance on multiple-choice comprehension test items. *Reading Research Quarterly*, 23, 454-464. <https://doi.org/10.2307/747643>
- Pressley M., Ghatala, E. S., Woloshyn, V., & Pirie, J. (1990). Sometimes adults miss the main ideas and do not realize it: Confidence in responses to short-answer and multiple-choice comprehension questions. *Reading Research Quarterly*, 25, 232-249. <https://doi.org/10.2307/748004>
- Ramsey, P. H., Ramsey, P. P., & Barnes, M. J. (1987). Effects of student confidence and item difficulty on test score gains due to answer changing. *Teaching of Psychology*, 14 (4), 206-210. [https://doi.org/10.1207/s15328023top1404\\_3](https://doi.org/10.1207/s15328023top1404_3)

- Reiling, E., & Taylor, R. (1972). A new approach to the problem of changing initial responses to multiple choice questions. *Journal of Educational Measurement*, 9, 67-70. <https://doi.org/10.1111/j.1745-3984.1972.tb00762.x>
- Richards, J. C., Platt, J., & Platt, H. (1992). *Longman dictionary of language teaching and applied linguistics*. Longman Group UK Limited.
- Riding, R., & Rayner, S. (1998). *Cognitive styles and learning strategies: Understanding style differences in learning and behavior*. David Fulton.
- Rozencwajg, P. & Corroyer, D. (2005). Cognitive processes in the reflective-impulsive cognitive style. *The Journal of Genetic Psychology: Research and Theory on Human Development*, 166(4), 451-463, <https://doi.org/10.3200/GNTP.166.4.451-466>
- Schwartz, S. P., McMorris, R. F., & DeMers, L. P. (1991). Reasons for changing answers: An evaluation using personal interviews. *Journal of Educational Measurement*, 28, 163–171.
- Sellah, L., Jacinta, K., Helen, M., & Qian, M. (2018). Predictive power of cognitive styles on academic performance of students in selected national secondary schools in Kenya. *Cogent Psychology*, 5(1). <https://doi.org/10.1080/23311908.2018.1444908>
- Soil, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 780-805. <https://doi.org/10.1037/a0015145>
- Stylianou-Georgiou, A., & Papanastasiou, E. C. (2017). Answer changing in testing situations: The role of metacognition in deciding which answers to review. *Educational Research and Evaluation*, 23(3-4), 102-118. <https://doi.org/10.1080/13803611.2017.1390479>
- Vispoel, W. P. (1998). Reviewing and changing answers on computer-adaptive and self-adaptive vocabulary tests. *Journal of Educational Measurement*, 35, 328–345.
- Vul, E., & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science*, 19, 645-647. <https://doi.org/10.1111/j.1467-9280.2008.02136.x>
- Yaniv, I., & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes*, 103, 104-120. <https://doi.org/10.1016/j.obhdp.2006.05.006>
- Zaidi, N., Grob, K., Monrad, S., Kurtz, J., Tai, A., Ahmed, A., Gruppen, L., & Santen, S. (2018). Pushing critical thinking skills with multiple-choice questions: Does Bloom's

Taxonomy work? Academic Medicine, 93 (6), 856-859,  
<https://doi.org/10.1097/ACM.0000000000002087>

Zelniker, T., & Jeffrey, W. E. (1976). Reflective and impulsive children: Strategies of information processing underlying differences in problem solving. Monographs of the Society for Research in Child Development, 41, 1–59. <https://doi.org/10.2307/1166020>

Zhang, J. (2023). Links between cognitive styles and learning strategies in second language acquisition. Journal of Education, Humanities and Social Sciences, 8, 327-333. <https://doi.org/10.54097/ehss.v8i.4269>



## Appendix A

### Regressions Coefficients for the Traditional Test

**Table A1.** Regression Coefficients for the Traditional Test (Changing to Wrong Options)

Model	Standardized Coefficients		t	Sig.
	Beta			
Field Dependent/Field Independent	.22		1.65	.11
Leveler/Sharpener	-.05		-.45	.65
Impulsive/Reflective	-.76		-5.72	<b>.00</b>
Global/Particular	.17		1.40	.17
Analog/Digital	-.05		-.44	.66
Synthetic/Analytic	-.10		-.86	.40
Concrete/Abstract	.08		.78	.44
Field Sensitive/Field Insensitive	.06		.58	.57
Random/Sequential	-.15		-1.31	.20
Inductive/Deductive	-.12		-1.16	.25

Dependent Variable: Changing to wrong options

**Table A2.** Regression Coefficients for the Traditional Test (Changing to Right Options)

Model	Standardized Coefficients		t	Sig.
	Beta			
Field Dependent/Field Independent	-.23		-1.76	.08
Leveler/Sharpener	-.06		-.56	.58
Impulsive/Reflective	-.58		-4.40	<b>.00</b>
Global/Particular	.09		.77	.45
Analog/Digital	-.02		-.17	.87
Synthetic/Analytic	.04		.38	.71
Concrete/Abstract	-.03		-.33	.74
Field Sensitive/Field Insensitive	-.05		-.43	.67
Random/Sequential	.02		.21	.84
Inductive/Deductive	-.16		-1.55	.13

Dependent Variable: Changing to right options

**Table A3.** *Regression Coefficients for the Traditional Test (Total Answer Changing)*

Model	Standardized Coefficients		t	Sig.
	Beta			
Field Dependent/Field Independent	-.11		-1.18	.25
Leveler/Sharpener	-.07		-.89	.38
Impulsive/Reflective	-.78		-8.15	<b>.00</b>
Global/Particular	.14		1.63	.11
Analog/Digital	-.04		-.43	.67
Synthetic/Analytic	-.00		-.01	.99
Concrete/Abstract	.00		.03	.98
Field Sensitive/Field Insensitive	-.02		-.19	.85
Random/Sequential	-.04		-.45	.66
Inductive/Deductive	-.18		-2.42	<b>.02</b>

Dependent Variable: Total answer changing

