Tabaran Institute of Higher Education
International Journal of Language Testing
Vol. 12, No. 1, March 2022

ISSN 2476-5880

Assessing Test-Taking Strategies of IELTS Test-Takers: Development and Validation of an IELTS Test-Taking Strategy Questionnaire

Masoomeh Estaji[1]*, Zahra Banitalebi[2]

Received: January 2022 Accepted: March 2022

Abstract

The measurement of test-taking strategies and practices, mostly studied through qualitative methods, has been an important aspect of language testing and assessment research. The current study examines the test-taking strategies of International English Language Testing System (IELTS) test-takers and reports the process of designing and validating an IELTS test-taking strategy questionnaire. To achieve this aim, a questionnaire with 62 items was developed and piloted on 534 IELTS test-takers. To ensure its validity, the questionnaire results were analyzed through Exploratory and Confirmatory Factor Analyses (EFA and CFA). The final 49-item instrument with eight factors, "test-management (TM)" and "test-wiseness (TW)" strategies in each skill, had adequate psychometric properties. The findings revealed positive correlations between TM and TW strategies, representing strong correlations between Reading TM and Listening TM, Reading TW and Listening TW, and Listening TM and Speaking TM. The developed questionnaire can serve as a diagnostic tool to monitor test-takers' performance and strategies, mainly in high-stakes tests like IELTS.

Keywords: Factor analysis; Questionnaire design; Test-management strategy; Test-wiseness strategy; Validation

1. Introduction

Despite their powerful impacts on language testing and teaching, tests, especially high-stakes ones, might happen to be limited in accuracy regarding the description of candidates' proficiency (e.g., Field, 2012). Language competence alone does not represent one's successful demonstration of ability and test performance (Aryadoust, 2019; Low & Aryadoust, 2021). In responding to the test items, candidates look for various strategies to boost their scores. Some of such strategies can be irrelevant to the construct of language competence (Cohen, 2006), resulting in false reports of test-takers' real language abilities. Cohen (2018) refers to two main reasons for conducting research on test-taking strategies. First, it can ensure whether or not tests measure what they intend to measure and produce evidence for validity of tests. Second, it is helpful to examine the true proficiency levels of language learners, which might be masked by clever responses on the part of test-takers. Hence, strategies, in particular test-taking

[1] Associate Professor of Applied Linguistics, Department of English Language and Literature, Faculty of Persian Literature and Foreign Languages, Allameh Tabataba'i University, Tehran, Iran.
Email: *mestaji74@gmail.com*
[2] Department of English Language and Literature, Faculty of Persian Literature and Foreign Languages, Allameh Tabataba'i University, Tehran, Iran. Email: *zahra.banitalebi@gmail.com*

ISSN 2476-5880

strategies (Aryadoust, 2019; Phakiti, 2008; Purpura, 1999), have been considered as an important aspect of test performance and validation (Anderson, Bachman, Perkins, & Cohen, 1991; Bachman, 1990; Cohen, 2006, 2012).

Tests need to demonstrate cognitive validity (Field, 2012; Glaser, 1991), meaning that the candidates should engage in a set of thinking processes that resemble procedures individuals employ in real-life events. The conception of cognitive validity and construction of valid items and tasks depend on the relevant cognitive processes testees utilize to answer the test (O'Sullivan & Weir, 2011) and give significance to the appropriateness of strategies test-takers mobilize to answer an item (Barati, Ravand, & Ghasemi, 2013). In a similar vein, Davidson (2000) has stressed the importance of test results interpretation based on evidence rather than pure statistics (e.g., correlation-based statistical analysis that prevails the field, Lim, 2020). Any inferences derived from the test scores might become the subject of a contentious debate if the testees' cognitive processes are disregarded (Ghahramanlou, Zohoorian, & Baghaei, 2017; Lim, 2020; Roohani Tonekaboni, Ravand, & Rezvani, 2021). Ideally, test performance should indicate the mastery of language components rather than outsmarting the test (Cohen, 2018). Therefore, examining test-taking strategies can serve as guidance to devise, develop, and validate language assessment measures.

Test-taking strategies generally refer to construct relevant and irrelevant (i.e., TM and TW) strategies that test-takers deploy to answer test items. While the former assists meaningful responses to the questions, the latter does not include any involvement of language ability or knowledge. Using TW strategies, testees may gain scores due to some construct-irrelevant factors, such as guessing (Cohen, 2012). Hence, if success at achieving scores on a test requires the utilization of TW strategies, the test can be rendered invalid. Given the centrality of test-taking strategies in validation assessment, many scholars have been inspired to work on this research domain (Cohen, 2012; Cohen & Upton, 2007). Tracing the existing literature, one can easily realize that a great majority of the studies have targeted the issue by assessing test-takers' verbal reports of their strategy uses. In other words, verbal reports have been the primary research tool in these studies (Cohen, 2012; Cohen & Upton, 2007).

Not only are questionnaires as data collection instruments rare but also the available questionnaires take no notice of TM strategies (Zhang, Goh, & Kunnan, 2014). The present-day account of the literature mostly describes the development of questionnaires, considering language learning strategies rather than test-specific strategies. Although TM and TW strategies are prominent in explaining the testees' performance, investigating test-taking strategies has not been given enough attention. More importantly, using quantitative methods to achieve this evidence is relatively rare (Wu & Zumbo, 2017).

General scrutiny of the published questionnaires for high-stake tests such as IELTS shows that there is no comprehensive questionnaire for measuring the testees' test-taking strategies. As IELTS has gained recognition from thousands of universities, schools, employers, and immigration bodies, millions of test-takers worldwide are striving to achieve acceptable IELTS band scores. However, satisfying this requirement is not an easy task unless test-takers deploy appropriate strategies to prepare for and take the IELTS exam. The term strategy might mistakenly refer to any good assistance (Cohen, 2018). In reality, however,

strategies can overestimate the testees' abilities and misplace individuals over their path of language learning, and in the case of IELTS lead to severe consequences for their academic and career lives. Hence, there is much more to understand about how test-taking strategies act and bring advantages to learners (Cohen, 2018). In this vein, Cohen (2018) calls for a closer look at test-taking strategies to resolve the complications.

In the absence of an all-inclusive test-taking strategy questionnaire, researchers have either developed their own questionnaires focusing only on one skill or resorted to other data gathering methods. Besides, research in this area needs to be taken more seriously to keep task and item designers on the right path (Cohen, 2012) and improve the experience of language learners (Cohen, 2018). To fill this gap, the present study examined the development and validation of a test-taking strategy questionnaire for IELTS, covering four skills of writing, speaking, listening, and reading.

2. Literature Review

Research on test-taking strategies can be divided into two main strands, corresponding to the two mentioned reasons for research on test-taking strategies given by Cohen (2018). One line of research has considered the advantages of test-taking strategies for validation purposes. Such practices follow Messick's lead for supporting construct validity. Messick (1995) refers to evidence by the test-takers' response processes to support the construct validity of the test. Dissatisfied with correlational-based methods to validity that cannot manifest the cognitive procedures, language testers have developed approaches to investigate test-taking strategies for validation purposes. They have utilized think-aloud protocols, eye-tracking analysis, and retrospective self-reports to dig into test-taking strategies. Understanding the relationship between test-taking strategies and test performance provides empirical evidence to examine if test-takers engage in relevant and appropriate strategies and skills, which adds to the construct validity of tests. The earliest interest in test-taking strategies refers back to as early as the 1980s, when the scholars started to probe the relationship between specific task types and the test-takers' strategy use (e.g., Cohen, 2006; Cohen & Upton, 2007) and examined the validity of different tests (e.g., TOEFL by Cohen & Upton, 2007). However, studying test-taking strategies for test validation has remained scarce so far (Lim, 2020).

In the other line of research, test-taking strategies have been incorporated into a broader concept of language learner strategies. It started with Rubin's (1975) seminal paper on learners' strategies in which the field of language learning shifted attention to what learners rather than teachers do to bring about success in language learning. New doors were opened to the field to explore what strategies successful learners employ and how they make them distinguishable from less successful learners. Since then, several studies have looked into learners' strategies in language learning and test-taking, establishing on the developed strategy inventories of O'Malley and Chamot (1990), Oxford (1990), Phakiti (2003), and Purpura (1999), among others. Although research on strategies has sparked growing interest among researchers (Zhang, Liu, Zhao, & Xie, 2011), the major focus of the previous studies was on language learning strategies that are consciously-held mental processes and behaviors that learners deploy when dealing with language-related aspects of the items (Cohen, 2011). Several

ISSN 2476-5880

scholars classified language learning strategies (O'Malley & Chamot, 1990; Oxford, 1990), many examined learners' learning and testing strategies relying on the participants' verbal accounts (Al Fraidan, 2019; Chappell, Yates, & Benson, 2019; Cohen & Upton, 2007; Fernandez, 2018), and some others developed strategy questionnaires (Phakiti, 2003, 2008; Purpura, 1999; Winke & Lim, 2017; Wu & Stone, 2016; Wu & Zumbo, 2017; Zhang, Aryadoust, & Zhang, 2013).

Generally, verbal reports have been the mostly-utilized methods for investigating test-takers strategies. For example, Al Fraidan (2019) used think-aloud protocols and interviews to examine learners' strategies to validate a cloze test and a multiple-choice vocabulary achievement test. In another study, Chappell et al. (2019) used interviews to report test-taking strategies that IELTS candidates deployed. Respondents were asked about test-taking strategies they knew. The results of structured interviews with 679 IELTS candidates reported a range of test-taking strategies, including both construct-irrelevant and construct-relevant strategies. The researchers compiled a list of strategies, classified into five categories. Strategies for reading, writing, listening, and speaking tasks were distinct from general strategies. Similarly, Barati (2005) proposed a test-taking strategy taxonomy focusing on reading English as a Foreign Language. This taxonomy included 41 items, classified under metacognitive and TW strategies.

Similarly, some researchers attempted to design questionnaires to gain insight into learners' use of strategies. Focusing on language learning strategies, Purpura (1999) devised a questionnaire to measure how learners would build up their language skills and vocabulary and grammar knowledge. Likewise, Phakiti (2003, 2008) developed reading strategy questionnaires. Although the devised instruments included items on TM strategies, they mainly relied on language learning strategies. In another study, taking TM strategies into account, Zhang et al. (2013) proposed and validated a strategy use questionnaire to measure the test-takers' metacognition in reading comprehension tests. Structural equation modeling and Rasch analysis provided validity evidence for their study.

More recently, Winke and Lim (2017) designed a series of questionnaires examining general listening and test-taking strategies to capture the influence of explicit test-taking instruction on IELTS listening test scores. For their test-taking strategy questionnaire, they adopted Cohen and Upton's (2007) questionnaire which was developed for the Internet-Based TOEFL® reading test. However, their questionnaire was modified and some items were added to cater for varied types of IELTS listening items. Finally, the questionnaire data went under the procedures of EFA and CFA for validation.

What this review signifies is that verbal reports are the most commonly used instruments in examining test-taking strategies. However, there are some advantages that questionnaires offer over verbal reports. While administering verbal reports requires extensive time, the data collection, administration, and interpretation of questionnaires are not much demanding (Zhang et al., 2013). Likewise, the generalizability of these verbal report methods is limited. Furthermore, a verbal report is a complex methodology surrounded by many misunderstandings and pitfalls (Green, 1998). That said, employing verbal reports requires some degree of orientation to avoid problems and enhance the study's quality (Green, 1998).

 ISSN 2476-5880

Despite the availability of questionnaires developed to investigate testees' strategies, they merely assessed test-takers' general learning strategies mainly in one or two skills. Moreover, given the significance of IELTS as a high-stakes test, a questionnaire with a specific focus on testing strategies is lacking. Driven by this dearth of research on test-takers' awareness and use of strategies and the shortage of a validated instrument to measure TM and TW strategies, the present study aimed to design and validate an IELTS test-taking strategy questionnaire.

3. Methodology

3.1. Participants

Five hundred and thirty-four responses were gathered from Iranian IELTS test-takers, who had recently taken the test in various cities in Iran. The participants were chosen following convenience sampling as the data were gathered from a conveniently available sample of IELTS candidates. Their age ranged from 19 to 38 (M= 25.95). Of all the respondents, 43.82% were identified as female and 56.18% male. A vast majority (74%) of the respondents sat an IELTS Academic test, while 26% took the General Training version. The participants were found homogeneous considering their overall IELTS scores (M= 5.8, SD= 0.5).

3.2. Instrument

The only instrument used in this study was the developed Test-Taking Strategy Questionnaire. The first section of the questionnaire asked for general information such as the participants' age, IELTS score, type of test, and gender, while in the second part they were required to self-report their strategies by marking their responses in the questionnaire.

3.3. Data Collection Procedure

The present study was carried out in two steps. In the initial phase, an online Test-Taking Strategy Questionnaire was developed. Questionnaire distribution took place with the assistance of IELTS Test Centers and IELTS academies in Iran. The online link of the questionnaire was posted on different Telegram channels created for testees. In the succeeding phase, the scale was validated in relation to the collected data from 534 Iranian IELTS test-takers (Appendix A for the final version of the questionnaire).

To develop the questionnaire, a rigorous review of the relevant literature in IELTS and test-taking strategies was performed. Following the standards and procedures of questionnaire development by Dörnyei and Taguchi (2010), the researchers also carried out interviews with experts of the field to enlarge the item pool. Using the survey items gathered from previous studies (Barati, 2005; Chappell et al., 2019; Cohen & Upton, 2007; Knoch, Huisman, Elder, Kong, & McKenna, 2020; Winke & Lim, 2017), a draft version of the questionnaire was developed. The work of Chappell et al. (2019) formed the basis for item generation because they had examined a large sample of IELTS test-takers (758), and their instruments focused on the testees' strategies in the four tasks. Therefore, their inventory was found both comprehensive and in harmony with the concerns of this study. To reach a more inclusive sample of items, extant instruments of other studies were considered as well. Relevant items from the mentioned studies were added; however, linguistically similar or repetitive survey

items were omitted. After assembling a representative sample, the items were organized so that each item could be rated from 1 (Not true of me at all) to 5 (Extremely true of me) on a 5-point Likert scale.

The next step was to optimize the content validity of the instrument. In doing so, four experts who had Ph.D. degrees, specializing in language assessment and applied linguistics, provided their feedback to ensure the content relevance and representativeness of the instrument. They used a color-coded scheme in selecting the items (i.e., green, red, and yellow for the items to retain, delete, and leave undecided respectively). Finally, 62 items judged to be suitable by the experts were retained.

Subsequently, to improve the quality and format of the survey (Creswell, 2009), the instrument was pilot-tested on a small sample taken from the target population. To this aim, 45 IELTS-takers were asked to participate in the piloting phase and give their comments about the instrument to check the clarity and readability of the items. The data collection took place online. After finalizing the questionnaire and ensuring its content validity, it was administered to 534 main participants of the study and went through the subsequent statistical analyses.

3.4. Data Analysis

To determine the construct validity of the instrument, EFA and CFA were performed. The original questionnaire had 62 items. Before commencing the analysis, the data went through some pre-processes to exclude the problematic data. Initially, 534 solid answers were obtained. No missing answer was found in the data. Then the standard deviation of the respondents' answers was calculated, and no case was found suspicious (standard deviation below 0.5) for an unengaged response. Finally, the item-total statistics were inspected for the reliability analysis of eight potential factors of the questionnaire (i.e., TW and TM in four language skills). This process resulted in excluding some items, specifically, Q5, Q9, and Q10 from *Writing TM*, Q24 from *Reading TM*, Q30 from *Reading TW*, Q34 and Q49 from *Listening TM*, Q50 and Q51 from *Listening TW*, and Q56 from *Speaking TM*. The deletion of these items either made no change in the overall reliability or improved it (see Appendix B for the details). Besides, Q28 and Q36 were excluded as they had loadings values below .4 to their correspondent factor (Table 2, below). Finally, Q6 was also excluded for having a low standardized estimate (Table 4, below). The final questionnaire, thus, was left with 49 items.

4. Results

4.1. Exploratory Factor Analysis

To validate the questionnaire, first, a factor analysis with maximum likelihood (ML) extraction method and Promax rotation was run. Fifty-two items (the ones that remained after reliability analysis) entered the factor analysis. The sample adequacy was checked through Kaiser-Meyer-Olkin (KMO) and Bartlett's tests. The results of the sample adequacy test (KMO = .942 > .6) suggested adequacy for the sample size. Bartlett's test result was also found statistically significant at $p < .01$, meaning the null hypothesis that the factors in the matrix are not independent of each other (an identity matrix) could be rejected.

ISSN 2476-5880

The factor analysis results demonstrated the extraction of 8 factors with 70.55% of the total variance explained (Table 1). The scree plot of the loadings (Figure 1) also shows that while there were some fluctuations, after the final break and the eight factor, the loadings were decreasing steadily.

Table 1.
Total Variance Explained by Each Factor

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	18.49	35.56	35.56	17.75	34.14	34.14
2	5.15	9.91	45.47	4.46	8.58	42.72
3	4.40	8.47	53.94	3.82	7.35	50.07
4	3.18	6.13	60.07	2.67	5.14	55.21
5	2.36	4.55	64.63	2.80	5.38	60.60
6	2.08	4.01	68.64	1.66	3.20	63.81
7	1.74	3.36	72.00	2.25	4.32	68.14
8	1.43	2.75	74.76	1.25	2.41	70.55
9	.93	1.79	76.55			
.	.	.	.			
.	.	.	.			
.	.	.	.			
52	.026	.051	100.000			

Extraction Method: Maximum Likelihood.

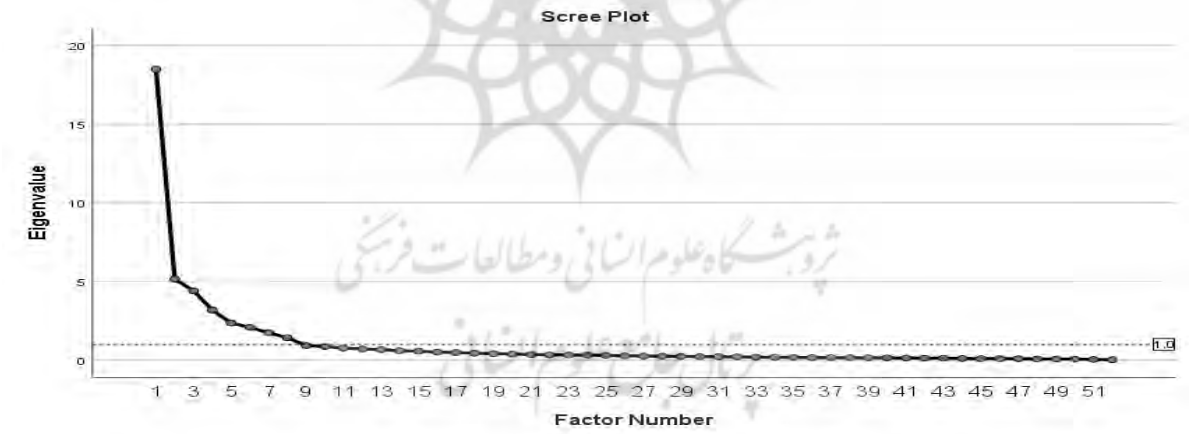

Figure 1. EFA Scree Plot

The pattern matrix resulted from this extraction is shown in Table 2.

Table 2.
EFA Pattern Matrix

	Factor							
	1	2	3	4	5	6	7	8
Q01							.78	
Q02							.87	
Q03							.86	
Q04						.91		

ISSN 2476-5880

	Factor							
	1	2	3	4	5	6	7	8
Q06							.55	
Q07						.96		
Q08						.94		
Q11		.74						
Q12		.92						
Q13		.44						
Q14		.87						
Q15		.82						
Q16		.87						
Q17				.73				
Q18				.87				
Q19				.88				
Q20		.68						
Q21				.95				
Q22				.86				
Q23				.74				
Q25		.73						
Q26		.60						
Q27		.86						
Q28								
Q29				.59				
Q31	.95							
Q32	.84							
Q33	.75							
Q35	.87							
Q36								
Q37	.69							
Q38	.77							
Q39	.71							
Q40			.90					
Q41			.94					
Q42			.89					
Q43			.90					
Q44			.94					
Q45	.99							
Q46	.98							
Q47			.80					
Q48	.85							
Q52					.69			
Q53					.71			
Q54					.64			
Q55					.82			
Q57					.77			
Q58								.73
Q59					.87			
Q60								.87
Q61								.93
Q62					.76			

Rotation converged in 7 iterations.

In Table 1, items with loadings below 0.4 are suppressed. Based on the questionnaire content, the first factor was named *Listening TM*, factor 2 *Reading TM*, factor 3 *Reading TW*, factor 4 *Listening TW*, factor 5 *Speaking TM*, factor 6 *Writing TW*, factor 7 *Speaking TW*, and factor 8 *Writing TM*. Two items (Q28 and Q36), which had loadings below 0.4 were excluded.

 ISSN 2476-5880

4.2. Confirmatory Factor Analysis

A CFA was run through IBM AMOS (version 24) based on the pattern obtained from EFA. Before running CFA, the questionnaire items were checked for normality by inspecting the skewness and kurtosis ratios (Table 3). As reported, the skewness and kurtosis values for all items were between the range of -3 to +3, which indicates the normality of distributions.

Table 3.
Descriptive Statistics for the Questionnaire Items

	N	Min.	Max.	Mean	SD	Skewness	Kurtosis
Q01	534	1	5	3.43	.92	-.33	-.09
Q02	534	1	5	3.41	.95	-.31	-.16
Q03	534	1	5	3.46	.92	-.37	-.04
Q04	534	1	5	2.97	1.48	-.00	-1.35
Q06	534	1	5	3.54	.96	-.43	-.06
Q07	534	1	5	2.99	1.56	-.02	-1.48
Q08	534	1	5	3.03	1.61	-.05	-1.55
Q11	534	1	5	3.18	1.17	-.14	-.78
Q12	534	1	5	2.82	1.08	.08	-.54
Q13	534	1	5	3.53	1.15	-.44	-.73
Q14	534	1	5	2.88	1.02	.06	-.47
Q15	534	1	5	3.10	1.00	.08	-.66
Q16	534	1	5	2.82	1.01	.15	-.33
Q17	534	1	5	2.66	1.10	.27	-.57
Q18	534	1	5	2.58	1.06	.22	-.55
Q19	534	1	5	2.55	1.09	.39	-.47
Q20	534	1	5	2.81	1.14	.34	-.60
Q21	534	1	5	2.59	1.20	.32	-.73
Q22	534	1	5	2.53	1.10	.42	-.39
Q23	534	1	5	2.61	1.13	.17	-.77
Q25	534	1	5	3.41	1.09	-.19	-.81
Q26	534	1	5	3.35	1.07	-.26	-.54
Q28	534	1	5	2.99	.98	.12	-.21
Q29	534	1	4	2.03	.81	.11	-1.10
Q31	534	1	5	3.08	1.09	.07	-.60
Q32	534	1	5	3.15	1.12	-.15	-.65
Q33	534	1	5	2.88	1.12	.26	-.61
Q35	534	1	5	3.06	1.11	-.01	-.52
Q37	534	1	5	3.24	1.15	-.17	-.70
Q38	534	1	5	2.91	1.07	.26	-.49
Q39	534	1	5	3.02	1.17	.04	-.76
Q40	534	1	5	2.57	1.27	.51	-.81
Q41	534	1	5	2.71	1.28	.41	-.92
Q42	534	1	5	2.84	1.28	.37	-.92
Q43	534	1	5	2.61	1.25	.43	-.87
Q44	534	1	5	2.61	1.29	.49	-.86
Q45	534	1	5	3.04	1.05	-.03	-.41
Q46	534	1	5	2.96	1.05	.08	-.36
Q47	534	1	5	2.89	1.19	.34	-.90
Q48	534	1	5	3.01	1.18	-.07	-.78
Q52	534	1	5	2.68	1.15	.21	-.75
Q53	534	1	5	2.40	1.04	.38	-.39
Q54	534	1	5	2.66	1.10	.38	-.31
Q55	534	1	5	2.39	.98	.31	-.34
Q57	534	1	5	2.35	.85	.04	-.58
Q58	534	1	4	2.06	.74	.40	-.01

ISSN 2476-5880

	N	Min.	Max.	Mean	SD	Skewness	Kurtosis
Q59	534	1	5	2.38	.99	.36	-.24
Q60	534	1	4	1.84	.74	.54	-.18
Q61	534	1	4	1.91	.72	.51	.13
Q62	534	1	5	2.58	.96	.24	-.03

Having ensured the normality of distributions, the CFA was run using IBM AMOS. To do so, first, items with non-significant loadings in unstandardized estimation were excluded. Table 4 elucidates the results for the standardized and unstandardized estimates. As reported, none of the items had a non-significant unstandardized estimate; however, Q06 was excluded as it had a standardized estimate below 0.5.

Table 4.
Standardized and Unstandardized Estimates of the Initial CFA Model

			Unstandardized				Standardized
			Estimate	S.E.	C.R.	P	Estimate
Q31	<---	Listening.TM	1.00				.81
Q32	<---	Listening.TM	1.04	.04	23.24	.00	.82
Q33	<---	Listening.TM	1.07	.04	24.47	.00	.85
Q35	<---	Listening.TM	1.05	.04	24.13	.00	.84
Q37	<---	Listening.TM	.96	.04	19.99	.00	.74
Q38	<---	Listening.TM	.99	.04	23.17	.00	.82
Q39	<---	Listening.TM	1.11	.04	24.27	.00	.85
Q45	<---	Listening.TM	1.14	.03	30.49	.00	.97
Q46	<---	Listening.TM	1.15	.03	30.61	.00	.97
Q48	<---	Listening.TM	1.13	.04	24.51	.00	.85
Q11	<---	Reading.TM	1.00				.74
Q12	<---	Reading.TM	1.03	.05	20.22	.00	.83
Q13	<---	Reading.TM	.75	.05	13.51	.00	.57
Q14	<---	Reading.TM	1.00	.04	20.79	.00	.85
Q15	<---	Reading.TM	.95	.04	20.26	.00	.83
Q16	<---	Reading.TM	1.05	.04	22.16	.00	.90
Q20	<---	Reading.TM	1.09	.05	20.23	.00	.83
Q25	<---	Reading.TM	.78	.05	15.10	.00	.64
Q26	<---	Reading.TM	.78	.05	15.12	.00	.64
Q27	<---	Reading.TM	.99	.05	19.67	.00	.81
Q17	<---	Reading.TW	1.00				.70
Q18	<---	Reading.TW	1.12	.06	18.59	.00	.82
Q19	<---	Reading.TW	1.18	.06	19.02	.00	.84
Q21	<---	Reading.TW	1.41	.06	20.57	.00	.91
Q22	<---	Reading.TW	1.27	.06	20.20	.00	.89
Q23	<---	Reading.TW	1.25	.06	19.44	.00	.86
Q29	<---	Reading.TW	.71	.04	15.47	.00	.68
Q40	<---	Listening.TW	1.00				.91
Q41	<---	Listening.TW	1.04	.02	39.12	.00	.94
Q42	<---	Listening.TW	1.02	.02	37.55	.00	.92
Q43	<---	Listening.TW	.99	.02	36.09	.00	.91
Q44	<---	Listening.TW	1.05	.02	39.91	.00	.94
Q47	<---	Listening.TW	.83	.03	26.26	.00	.80
Q52	<---	Speaking.TM	1.00				.74
Q53	<---	Speaking.TM	.95	.05	17.99	.00	.78
Q54	<---	Speaking.TM	.87	.05	15.50	.00	.68
Q55	<---	Speaking.TM	.90	.05	17.98	.00	.78
Q57	<---	Speaking.TM	.73	.04	16.93	.00	.74
Q59	<---	Speaking.TM	.92	.05	18.17	.00	.79
Q62	<---	Speaking.TM	.87	.04	17.97	.00	.78
Q04	<---	Writing.TW	1.00				.95

 ISSN 2476-5880

			Unstandardized				Standardized
			Estimate	S.E.	C.R.	P	Estimate
Q07	<---	Writing.TW	1.03	.01	55.91	.00	.96
Q08	<---	Writing.TW	1.05	.02	53.58	.00	.96
Q58	<---	Speaking.TW	1.00				.83
Q60	<---	Speaking.TW	1.04	.04	24.90	.00	.87
Q61	<---	Speaking.TW	1.07	.04	26.57	.00	.92
Q01	<---	Writing.TM	1.00				.92
Q02	<---	Writing.TM	.84	.03	21.62	.00	.75
Q03	<---	Writing.TM	.92	.03	27.06	.00	.85
Q06	<---	Writing.TM	.43	.05	8.67	.00	.37

The next step was to check the reliability and validity of the model. Before doing the analyses, those modifications proposed by AMOS to improve the model fit were applied. To do so, modifications with the threshold of 10 which were not contradictory to the literature were considered. Figure 1 delineates the final modified CFA model.

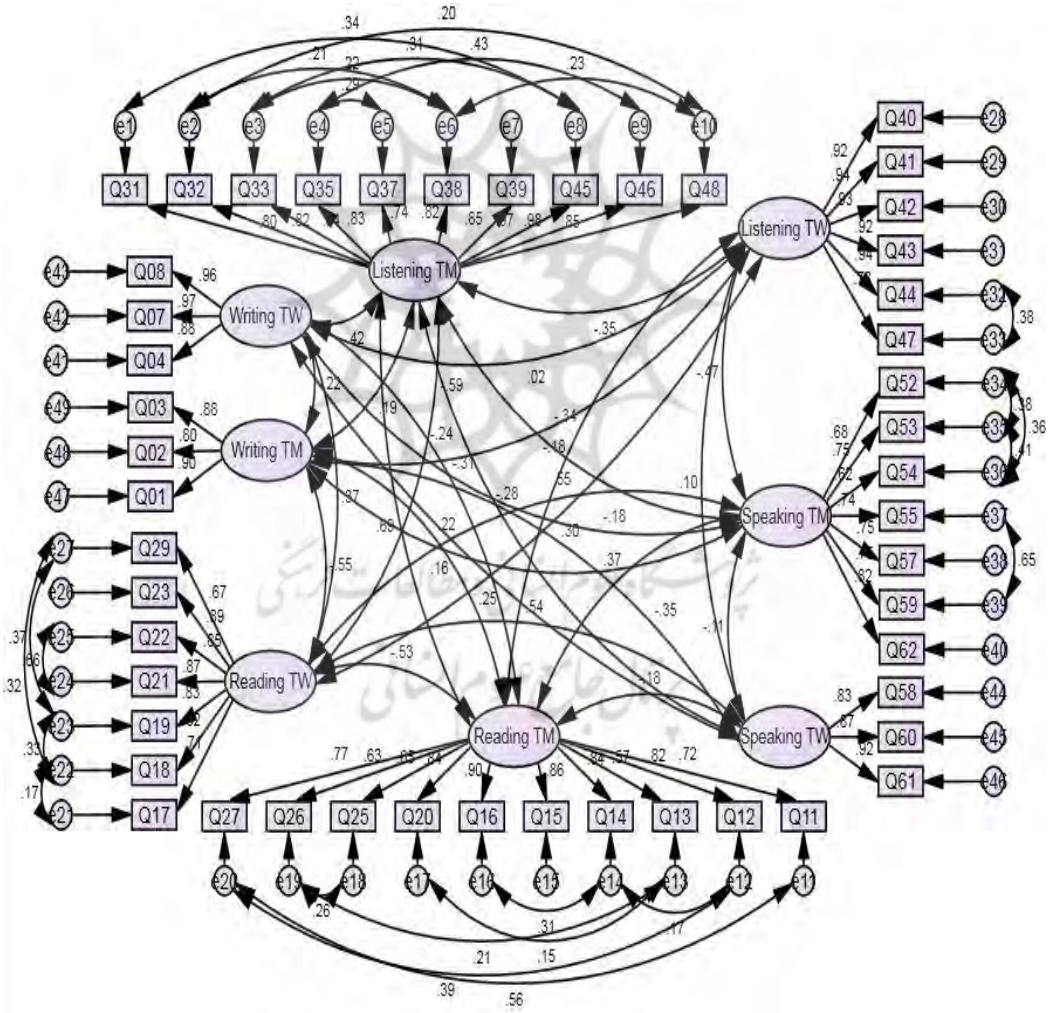

Figure 2. The Final Modified CFA Model with Standardized Estimates

After applying the modifications, the model's goodness of fit was examined. According to Hu and Bentler (1999), for the model to have acceptable goodness of fit, some criteria have to be met. These criteria, alongside the values obtained from the data, are reported (Table 5).

ISSN 2476-5880

Table 5.
Evaluation of the CFA Goodness of Fit

Criteria		Threshold			Evaluation
		Terrible	Acceptable	Excellent	
CMIN	3139.78				
df	1074				
CMIN/df	2.92	> 5	> 3	> 1	Excellent
RMSEA	.06	> 0.08	< 0.08	< 0.06	Acceptable
CFI	.92	< 0.9	> 0.9	> 0.95	Acceptable
TLI	.91	< 0.9	> 0.9	> 0.95	Acceptable
SRMR	.06	> 0.1	> 0.08	< 0.08	Excellent

CFI: Comparative Fit Index
CMIN: Minimum discrepancy function of C (Chi-Square)
df: degree of freedom
RMSEA: Root Mean Square Error of Approximation
SRMR: Standardized Root Mean Squared Residual

The results reported in Table 5 indicate acceptable to excellent goodness of fit. Next, the composite reliability (CR) and discriminant validity for each factor were examined (Table 6). As reported, all of the variables had CR values above 0.7, which reveals acceptable reliability. Besides, the average variance explained (AVE) values were both above 0.5 and lower than CR, suggesting convergent validity. Moreover, the square roots of AVE (the bold values in Table 6) for each factor were above their inter-correlations with other factors, implying discriminant validity according to Fornell and Larcker's (1981) criteria.

Table 6.
Composite Reliability and Discriminant Validity of the Factors

	CR	AVE	Fornell-Larcker Criterion							
			F1	F2	F3	F4	F5	F6	F7	F8
F1	0.96	0.72	**0.85**							
F2	0.93	0.59	0.68**	**0.76**						
F3	0.93	0.65	-0.58**	-0.53**	**0.81**					
F4	0.96	0.82	-0.35**	-0.34**	0.30**	**0.90**				
F5	0.89	0.53	0.54**	0.37**	-0.27*	-0.46**	**0.73**			
F6	0.95	0.88	-0.42**	-0.31**	0.37**	0.01	-0.17**	**0.93**		
F7	0.91	0.77	-0.24**	-0.18**	0.54**	0.09*	-0.11*	0.24**	**0.87**	
F8	0.89	0.73	0.19**	0.21**	-0.55**	-0.17**	0.15**	-0.21**	-0.35**	**0.86**

* Correlation is significant at $p < .05$
** Correlation is significant at $p < .01$
F1: Listening TM; F2: Reading TM; F3: Reading TW; F4: Listening TW; F5: Speaking TM; F6: Writing TW; F7: Speaking TW; F8: Writing TM

The inspection of the correlations (values not in bold under Fornell-Larcker's Criterion) documented that there are significant and positive correlations between all pairs of skills that measure TM (i.e., F1, F2, F5, and F8) as well as all pairs of skills that measure TW (i.e., F3, F4, F6, and F7), except for the correlation between listening TW and writing TW. The correlations between the pairs across TM and TW for all skills were negatively significant, indicating the existence of a trade-off between the two types of strategies. Moreover, the magnitude of negative correlations across the two types of strategies was small to medium (-

0.113 to -0.42) in most of the cases where two pairs of correlations showed strong values, i.e., Reading TW and Reading TM (r = 0.531); and Reading TW and Listening TM (r = -0.583).

Among the positive correlations in each type of strategies, there were also strong correlations between Reading TM and Listening TM (r = 0.687) and between Listening TM and Speaking TM (r = 0.547). These results indicate that while strong correlations may exist under TM strategies used for different skills, the correlations among the TW strategies used across the language skills are not as strong. Moreover, the strong correlations between the strategies used in reading and listening considering both TW (positive correlation) and across TW and TM (negative) indicate the closeness of strategies used for these two skills in both categories.

5. Discussion

The present study set out to design and validate an IELTS test-taking strategy questionnaire. It mainly attempted to bridge the gap in testing research, particularly in close relation to test-taking strategy assessment, by proposing and validating an all-inclusive IELTS Test-Taking Strategy Questionnaire. EFA and CFA were used to analyze the data. Initially, the Bartlett and KMO tests were applied to check and approve the sample adequacy. Then an EFA was run to detect the chief factors and condense the data. Through EFA, eight factors were extracted explaining 70.55% of the total variances. The model was further submitted to CFA, where some items were eliminated. After omitting some items, the results indicated that the questionnaire met adequate reliability and validity criteria for use in assessment, research, and educational contexts.

The final questionnaire contains 49 items with eight factors examining TM and TW in writing, speaking, listening, and reading skills. The constituent structure of the present study's model is consistent with Cohen's proposals (Cohen, 2006, 2007, 2012; Cohen & Upton, 2007) and previous studies (Barati, 2005; Chappell et al., 2019; Winke & Lim, 2017). However, unlike the studies of Barati (2005), Phakiti (2003, 2008), Wu and Stone (2016), and Wu and Zumbo (2017), to have adequate items representing each factor the subcomponents of TM strategies were not identified.

TM strategies in this study are mainly considered as a general term referring to construct-relevant strategies. However, in Wu and Stone's (2016) study, three factors, namely, TM, TW, and comprehending strategies were identified as the testing strategies by the relevant factor analyses on the Canadian English Language Proficiency Index Program-General data. One justification is that a consolidated and integrated model for test-taking strategies has to emerge (Cohen, 2006; Wu & Stone, 2016). Therefore, researchers have taken up differing positions in distinguishing between different types of testing strategies or including learning strategies in test-taking strategy typologies (Bumbálková, 2021).

These findings primarily support this assumption that test-taking strategies chiefly comprise TM and TW strategies by indicating positive correlations between the pairs of skills measuring TM and those measuring TW. In support of uniting strategies rather than differentiating them in detail, the results of Nikolov's (2006) study unraveled that attempts to subcategorize strategies might happen to be in vain due to the likelihood of strategy overlap.

Thus, extensive sub-division of test-taking strategies seems problematic as many factors would have been involved.

Among the positive correlations between TM and TW strategies, the findings revealed strong correlations between Reading TM and Listening TM, Reading TW and Listening TW, and Listening TM and Speaking TM. This is in line with Cohen's (2007) and Nikolov's (2006) findings that some testing strategies are highly likely to co-occur in clusters. In the case of Listening and Reading, due to the nature of the majority of IELTS Listening and Reading test types, comprising multiple-choice items, several items in the questionnaire in both parts might address the testee's tendency toward answering the specific type of questions with or without regard to a particular skill.

Overall, the correlations between the pairs across TM and TW for each skill were negative, pinpointing the natural existence of a trade-off between these two types of strategies. Although the magnitude of correlations was small to medium, this is at odds with Yang's (2000) findings suggesting that test-wise students have better TM strategies in the listening and reading subtests of paper-and-pencil-based TOEFL. One justification is that TOEFL has undergone substantial modification to improve its validity. As a consequence, items may not be much susceptible to TW strategies as they were before. The outgrowth of TW and washback studies have positively impacted the test developers to consider the interconnections between the test results, program practices, and individual learner strategies (Green, 2013). In other studies, Wu and Stone (2016) and Wu and Zumbo (2017), despite having a different classification of strategies, also found a negative association between test-wiseness strategies and test performance, meaning that the fewer testees utilized this type of strategies, the better results they gained.

All in all, the extracted components indicate that TM and TW are significant facets of testees' test-taking strategies. The good model fit of the study's CFA provided additional evidence for the validity aspects, indicating that the latent constructs of TM and TW were well-operationalized by the questionnaire.

6. Conclusion and Implications

This study sought to design and validate an IELTS test-taking strategy questionnaire. Based on the outcomes, it can be concluded that the construct of test-taking strategy has nested components in the IELTS examination. As the test-taking strategy is constructed through the companionship of TM and TW, these two have an interrelationship with each other, especially in the IELTS test due to the nature and type of subtests.

This study holds multiple implications for the field. Theoretically, this study can extend the literature on test-specific strategies which happen to be remained under the shadow of general language learning strategies to date (Bumbálková, 2021). While qualitative research more effectively identifies varied strategies adopted by testees, it might not be as efficient for unraveling the network of relationships among strategies. By developing and validating a test-taking strategy questionnaire of a high-stakes international test, this study can pave the way to further quantitative research in this area. Avid researchers can build up on the findings of the present study and advance investigations on the complex relationship between TM and TW

strategies. Since testing processes have been a valued source of understanding validity, the developed questionnaire can also be utilized to provide validity evidence in future studies.

This study also offers practical implications for IELTS developers and users such as teachers, learners, test designers, and educational researchers. For instance, for IELTS instructors the developed questionnaire can act as a diagnostic tool to monitor test-takers' performance on tests. Previous studies (e.g., Ata, 2015; Burton, 2020; Rasti, 2009) have shown that misconceptions about IELTS test items, the required strategies, and language learning, in general, can hinder students' progress in acquiring language skills. Therefore, teachers can inform the candidates about the effectiveness of strategies they use, advise them how to properly use these strategies based on the task requirements in each skill, and help them resolve the probable misconceptions about IELTS. Likewise, teachers can employ test-taking questionnaires to assess the quality of IELTS training courses concerning the positive or negative washback that their instructions, materials, and textbooks may have, particularly on testees' test-taking strategies. Additionally, as suggested by Cohen (2012), test-wiseness studies need to continue "as a means of checking whether tests are giving away the answers to items" (p. 267). Thus, this path of studies should be pursued to inform test and item developers.

Although the participants of this study were large enough to represent the general characteristics of the IELTS test-takers, this may still threaten the generalizability of the results. Researchers have suggested that test-takers' perceptions and performance can be affected by context-specific political, economic, and educational realities. This study would be strengthened if a multi-contextual approach could have been adopted. Finally, the accuracy and reliability of self-reports in describing testees' cognitive processes have been a point under discussion for a long time. For instance, compared with self-report questionnaires, the use of the eye-tracking method seems to be a stronger predictor of test-taking strategy use (Bax & Weir, 2012; Low & Aryadoust, 2021). Nonetheless, the feasibility of this instrument, in comparison with other methods, has made it a common use in the field.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Al Fraidan, A. (2019). New test-taking patterns and their effect on language test validity. *Language Testing in Asia*, *9*(1), 1-28. https://doi.org/10.1186/s40468-019-0088-5

Anderson, N. J., Bachman, L., Perkins, K., & Cohen, A. D. (1991). An exploratory study into the construct validity of a reading comprehension test: Triangulation of data sources. *Language Testing*, *8*(1), 41-66. https://doi.org/10.1177/026553229100800104

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*, *28*(6), 550-576. https://doi.org/10.1080/09588221. 2019.1574267

Ata, A. W. (2015). Knowledge, education, and attitudes of international students to IELTS: A case of Australia. *Journal of International Students, 5*(4), 488-500. http://dx.doi.org/10.32674/jis.v5i4.410

Bachman, L. F. (1990). *Fundamental consideration in language testing.* Oxford, UK: Oxford University Press.

Barati, H. (2005). *Test-taking strategies and the assessment of reading skills: An approach to construct validation* (Doctoral dissertation). University of Bristol.

Barati, H., Ravand, H., & Ghasemi, V. (2013). Investigating relationships among test takers' characteristics and response formats in a reading comprehension test: A structural equation modeling approach. *International Journal of Language Testing*, *3*(2), 38-59.

Bax, S., & Weir, C. (2012). Investigating learners' cognitive reading processes during a computer-based CAE reading test. *Research Note, 47,* 3-14. https://hdl.handle.net/10547/337904

Bumbálková, E. (2021). Test-taking strategies in second language receptive skills tests: A literature review. *International Journal of Instruction*, *14*(2), 647-664. https://doi.org/10.29333/iji.2021.14236a

Burton, J. D. (2020). How scripted is this going to be? Raters' views of authenticity in speaking-performance tests. *Language Assessment Quarterly*, *17*(3), 244-261. https://doi.org/10.1080/15434303.2020.1754829

Chappell, P., Yates, L., & Benson, P. (2019). *Investigating test preparation practices: Reducing risks.* (IELTS Research Reports Online Series, No. 3). https://www.ielts.org/-/media/research-reports/20193-chappell_et_al_layout.ashx

Cohen, A. D. (2006). The coming of age of research on test-taking strategies. *Language Assessment Quarterly*, *3*(4), 307-331. https://doi.org/10.1080/15434300701333129

Cohen, A. D. (2007). Coming to terms with language learner strategies: Surveying the experts. In A. D. Cohen & E. Macaro (Eds.), *Language Learner Strategies: 30 years of Research and Practice* (pp. 29-45). Oxford: Oxford University Press.

Cohen, A. D. (2011). Test-taking strategies. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoynoff (Eds.), *The Cambridge guide to second language assessment* (pp. 96-104). Cambridge: Cambridge University Press.

Cohen, A. D. (2012). Test taker strategies and task design. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 262-277). London and New York: Routledge.

Cohen, A. D. (2018). Reflections on a career in second language studies: Promising pathways for future research. *L2 Journal*, *10*(1), 1-19. https://doi.org/10.5070/L210136653

Cohen, A. D., & Upton, T. A. (2007). I want to go back to the text: Response strategies on the reading subtest of the new TOEFL®. *Language Testing*, *24*(2), 209-250. https://doi.org/10.1177/0265532207076364

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Los Angeles: Sage Publication.

Davidson, F. (2000). The language tester's statistical toolbox. *System*, *28*(4), 605-617. https://doi.org/10.1016/s0346-251x(00)00041-5

Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.). New York: Routledge.

Fernandez, C. J. (2018). Behind a spoken performance: Test takers' strategic reactions in a simulated part 3 of the IELTS speaking test. *Language Testing in Asia, 8*(18), 1-20. https://doi.org/10.1186/s40468-018-0073-4

Field, J. (2012). The cognitive validity of the lecture listening section of the IELTS listening paper. In L. Taylor & C. Weir (Eds.), *IELTS collected papers 2: Research in reading and listening assessment* (pp. 391-453). Cambridge: Cambridge University Press.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research, 18*(1), 39-50. https://doi.org/10.1177/002224378101800104

Ghahramanlou, M., Zohoorian, Z., & Baghaei, P. (2017). Understanding the cognitive processes underlying performance in the IELTS listening comprehension test. *International Journal of Language Testing*, *7*(1), 62-72. https://doi.org/10.20944/preprints201608.0190.v1

Glaser, R. (1991). Expertise and assessment. In M. C. Wittrock & E. L. Baker (Eds.), *Testing and cognition* (pp. 17-30). Englewood Cliffs, NJ: Prentice-Hall.

Green, A. (1998). *Using verbal protocols in language testing research: A handbook*. Cambridge: Cambridge University Press.

Green, A. (2013). Washback in language assessment. *International Journal of English Studies, 13*(2), 39-51. https://doi.org/10.6018/ijes.13.2.185891

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Knoch, U., Huisman, A., Elder, C., Kong, X., & McKenna, A. (2020). Drawing on repeat test takers to study test preparation practices and their links to score gains. *Language Testing, 37*(4), 550-572. https://doi.org/10.1177/0265532220927407

Lim, H. (2020). Exploring the validity evidence of a high-stake, second language reading test: An eye-tracking study. *Language Testing in Asia, 10*(1), 1-29. https://doi.org/10.1186/s40468-020-00107-0

Low, A. R. L., & Aryadoust, V. (2021). Investigating test-taking strategies in listening assessment: A comparative study of eye-tracking and self-report questionnaires. *International Journal of Listening*, 1-20. https://doi.org/10.1080/10904018.2021.1883433

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749. http://dx.doi.org/10.1037/0003-066X.50.9.741

Nikolov, M. (2006). Test taking strategies of 12 and 13 year old Hungarian learners of EFL: Why whales have migraines. *Language Learning, 56*(1), 1-51. https://doi.org/10.1111/j.0023-8333.2006.00341.x

O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition.* Cambridge: Cambridge University Press.

O'Sullivan, B., & Weir, C. J. (2011). Test development and validation. In B. O'Sullivan (Ed.), *Language Testing: theories and practice* (pp. 13-32). Oxford: Palgrave Macmillan.

Oxford, R. L. (1990). *Language learning strategies: What every teacher should know.* Boston: Heinle & Heinle.

Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing, 20*(1), 26-56. http://dx.doi.org/10.1191/0265532203lt243oa

Phakiti, A. (2008). Construct validation of Bachman and Palmer's (1996) strategic competence model over time in EFL reading tests. *Language Testing, 25*(2), 237-272. http://dx.doi.org/10.1177/0265532207086783

Purpura, J. E. (1999). *Learner strategy use and performance on language tests: A structural equation modeling approach*. Cambridge: Cambridge University Press.

Rasti, I. (2009). Iranian candidates' attitudes towards IELTS. *Asian EFL Journal*, *11*(3), 110-155.

Roohani Tonekaboni, F., Ravand, H., & Rezvani, R. (2021). The construction and validation of a Q-matrix for a high-stakes reading comprehension test: A G-DINA study. *International Journal of Language Testing*, *11(*1), 58-87.

Rubin, J. (1975). What the "good language learner" can teach us. *TESOL Quarterly, 9*(1), 41–51. http://dx.doi.org/10.2307/3586011

Winke, P., & Lim, H. (2017). The effects of test preparation on second-language listening test performance. *Language Assessment Quarterly*, *14*(4), 380-397. https://doi.org/10.1080/15434303.2017.1399396

Wu, A. D., & Stone, J. E. (2016). Validation through understanding test-taking strategies: An illustration with the CELPIP-General reading pilot test using structural equation modeling. *Journal of Psychoeducational Assessment*, *34*(4), 362-379. https://doi.org/10.1177/0734282915608575

Wu, A. D., & Zumbo, B. D. (2017). Understanding test-taking strategies for a reading comprehension test via latent variable regression with Pratt's importance measures. In B. D. Zumbo & A. M. Hubley (Eds.), *Understanding and investigating response processes: Advances in validation research* (pp. 305-319). Cham, Switzerland: Springer.

Yang, P. (2000). *Effects of test-wiseness upon performance on the test of English as a Foreign Language* (Doctoral dissertation). University of Alberta, Edmonton, CN.

Zhang, L., Aryadoust, V., & Zhang, L. J. (2013). Development and validation of the test takers' metacognitive awareness reading questionnaire (TMARQ). *The Asia-Pacific Education Researcher*, *23*(1), 37-51. http://dx.doi.org/10.1007/s40299-013-0083-z

Zhang, L., Goh, C. C., & Kunnan, A. J. (2014). Analysis of test takers' metacognitive and cognitive strategy use and EFL reading test performance: A multi-sample SEM approach. *Language Assessment Quarterly*, *11*(1), 76-102. https://doi.org/10.1080/15434303.2013.853770

Zhang, W., Liu, M., Zhao, S., & Xie, Q. (2011). English test-taking strategy use and students' test performance. *Asian EFL Journal*, *13*(2), 133-168.

Tabaran Institute of Higher Education ISSN 2476-5880
International Journal of Language Testing
Vol. 12, No. 1, March 2022

Appendix A: IELTS Test-Taking Strategy Questionnaire

The objective of this questionnaire is to glean information about the various strategies you have employed when taking the IELTS tests.

Part 1

1. Please fill in the following information.

- Name: (Initials only)
- Age:
- Gender: female male

2. Which module of the IELTS test did you take? (Academic or General)

3. What band scores did you achieve?

Listening	Writing	Speaking	Reading	Overall

Part 2

After reading each statement, choose the number (1, 2, 3, 4, or 5) which reflects what you actually did. Note that there are no right or wrong responses to any of the items on this questionnaire.

	1: Not true of me at all	2: Slightly true of me	3: Moderately true of me	4: Very true of me	5: Extremely true of me

1. For the writing tasks of the IELTS test, I used this strategy: **Organize/structure text appropriately.**

2. For the writing tasks of the IELTS test, I used this strategy: **Complete task two first.**

3. For the writing tasks of the IELTS test, I used this strategy: **Manage time carefully for each section.**

4. For the writing tasks of the IELTS test, I used this strategy: **Know how to answer different question types.**

5. For the writing tasks of the IELTS test, I used this strategy: **Insert a lot of big words into my writing, disregarding their use and context.**

6. For the writing tasks of the IELTS test, I used this strategy: **Rather than being error-free, have a wide range of grammatical structures.**

7. For the reading tasks of the IELTS test, I used this strategy: **Use skimming and scanning.**

8. For the reading tasks of the IELTS test, I used this strategy: **Know how to answer different question types.**

9. For the reading tasks of the IELTS test, I used this strategy: **Match sections of the text with the test questions.**

10. For the reading tasks of the IELTS test, I used this strategy: **Manage stress by acknowledging that the whole text does not need to be understood.**

11. For the reading tasks of the IELTS test, I used this strategy: **Identify synonyms that match the meaning between the question and the text.**

12. For the reading tasks of the IELTS test, I used this strategy: **Use the first two sentences of a paragraph to identify its main idea.**

13. For the reading tasks of the IELTS test, I used this strategy: **Look for an option that seems to deviate from the others (it is special or different).**

14. For the reading tasks of the IELTS test, I used this strategy: **Select a choice that is longer or shorter than the others.**

15. For the reading tasks of the IELTS test, I used this strategy: **Take advantage of clues appearing in other items to answer the item under consideration.**

16. For the reading tasks of the IELTS test, I used this strategy: **Double-check my answer to see if it is not awkward in context.**

17. For the reading tasks of the IELTS test, I used this strategy: **Select an option even though it is not understood, out of a vague sense that the other options couldn't be correct.**

18. For the reading tasks of the IELTS test, I used this strategy: **Use clues in other items to answer an item under consideration.**

19. For the reading tasks of the IELTS test, I used this strategy: **Select the option because it appears to have a word or phrase from the passage in it – possibly a keyword.**

20. For the reading tasks of the IELTS test, I used this strategy: **Manage time carefully.**

21. For the reading tasks of the IELTS test, I used this strategy: **Read all questions before reading the texts.**

22. For the reading tasks of the IELTS test, I used this strategy: **Highlight keywords.**

23. For the reading tasks of the IELTS test, I used this strategy**: Take into consideration the position of the correct options among the choices (a, b, c, or d) to find a pattern.**

24. For the listening tasks of the IELTS test, I used this strategy: **Read questions first and predict answers.**

25. For the listening tasks of the IELTS test, I used this strategy: **Take notes while listening.**

26. For the listening tasks of the IELTS test, I used this strategy: **Highlight keywords.**

27. For the listening tasks of the IELTS test, I used this strategy: **Use gaps between texts to read ahead for keywords.**

28. For the listening tasks of the IELTS test, I used this strategy: **Spell accurately and check the spellings afterward.**

29. For the listening tasks of the IELTS test, I used this strategy: **Know that the listening text becomes more complex toward the end.**

30. For the listening tasks of the IELTS test, I used this strategy: **Eliminate incorrect options while listening.**

31. For the listening tasks of the IELTS test, I used this strategy: **Look for an option that seems to deviate from the others (it is special or different).**

32. For the listening tasks of the IELTS test, I used this strategy: **Select a choice that is longer or shorter than the others.**

33. For the listening tasks of the IELTS test, I used this strategy: **Take advantage of clues appearing in other items to answer the item under consideration.**

34. For the listening tasks of the IELTS test, I used this strategy: **Use clues in other items to answer an item under consideration.**

35. For the listening tasks of the IELTS test, I used this strategy: **Select the option because it appears to have a word or phrase from the passage in it – possibly a keyword.**

36. For the listening tasks of the IELTS test, I used this strategy: **Pay extra attention to measurement units.**

37. For the listening tasks of the IELTS test, I used this strategy: **Pay extra attention to numbers.**

38. For the listening tasks of the IELTS test, I used this strategy: **Eliminate options that appear to be overlapping.**

39. For the listening tasks of the IELTS test, I used this strategy: **Know how to answer different question types.**

40. For the speaking tasks of the IELTS test, I used this strategy: **Use a wide range of vocabulary.**

41. For the speaking tasks of the IELTS test, I used this strategy: **Manage my affective state: relax/be natural/be calm/be confident.**

42. For the speaking tasks of the IELTS test, I used this strategy: **Speak clearly/avoid pronunciation errors.**

43. For the speaking tasks of the IELTS test, I used this strategy: **Explain/elaborate.**

44. For the speaking tasks of the IELTS test, I used this strategy: **Ask the examiner for clarification.**

45. For the speaking tasks of the IELTS test, I used this strategy: **Use fillers (e.g., "that's an interesting question") to avoid pausing.**

46. For the speaking tasks of the IELTS test, I used this strategy: **Use past, present, and future tenses, and active and passive voice (in parts 2 and 3).**

47. For the speaking tasks of the IELTS test, I used this strategy: **Know how to answer different question types.**

48. For the speaking tasks of the IELTS test, I used this strategy: **Change the quality of my voice.**

49. For the speaking tasks of the IELTS test, I used this strategy: **Imitate just a British or American accent to do well.**

Appendix B: Item-Total Statistics for the Reliability of TW and TM in Four Language Skills

Item-Total Statistics for Writing TM

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q01	17.07	9.64	.57	.54
Q02	17.09	8.92	.69	.50
Q03	17.07	9.36	.62	.53
Q05	16.85	11.12	.20	.66
Q06	17.06	10.06	.42	.59
Q09	15.01	12.67	.14	.66
Q10	16.94	12.05	.02	.73

Reliability Statistics for Writing TM

	Cronbach's Alpha	N of Items
Initial	.65	7
After exclusion of problematic items	.83	4

ISSN 2476-5880

Item-Total Statistics for Writing TW

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q04	6.02	9.75	.94	.96
Q07	5.96	9.52	.94	.95
Q08	5.91	9.26	.94	.96

Reliability Statistics for Writing TW

Cronbach's Alpha	N of Items
.97	3

Item-Total Statistics for Reading TM

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q11	33.95	78.20	.74	.92
Q12	34.37	78.71	.78	.91
Q13	33.59	82.07	.56	.92
Q14	34.28	79.43	.78	.91
Q15	34.06	79.74	.79	.91
Q16	34.35	78.75	.83	.91
Q20	34.38	77.59	.79	.91
Q24	34.56	86.39	.39	.93
Q25	33.74	81.76	.62	.92
Q26	33.80	81.67	.63	.92
Q27	34.05	78.62	.80	.91
Q28	34.20	83.54	.58	.92

Reliability Statistics for Reading TM

	Cronbach's Alpha	N of Items
Initial	.92	12
After exclusion of problematic items	.93	11

Item-Total Statistics for Reading TW

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q17	17.72	26.58	.69	.81
Q18	17.80	25.66	.82	.79
Q19	17.83	25.41	.82	.79
Q21	17.78	24.23	.84	.78
Q22	17.85	25.47	.80	.79
Q23	17.77	25.38	.78	.79
Q29	18.35	29.29	.65	.82
Q30	17.53	41.46	-.50	.93

Reliability Statistics for Reading TW

	Cronbach's Alpha	N of Items
Initial	.84	8
After exclusion of problematic items	.93	7

Item-Total Statistics for Listening TM

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q31	38.22	108.35	.78	.93
Q32	38.15	107.10	.81	.93
Q33	38.42	107.73	.79	.93
Q34	37.70	111.37	.51	.94
Q35	38.24	107.30	.82	.93
Q36	37.78	108.01	.72	.93
Q37	38.06	107.84	.76	.93

ISSN 2476-5880

Q38	38.40	108.46	.79	.93
Q39	38.28	106.19	.82	.93
Q45	38.27	106.18	.92	.92
Q46	38.34	105.93	.93	.92
Q48	38.30	105.84	.83	.93
Q49	37.46	128.59	-.09	.96

Reliability Statistics for Listening TM

	Cronbach's Alpha	N of Items
Initial	.94	13
After exclusion of problematic items	.96	11

Item-Total Statistics for Listening TW

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q40	18.78	36.25	.87	.84
Q41	18.65	35.71	.90	.83
Q42	18.52	35.93	.88	.84
Q43	18.75	36.33	.87	.84
Q44	18.74	35.45	.91	.83
Q47	18.46	38.33	.77	.85
Q50	19.07	44.21	.46	.88
Q51	18.53	56.85	-.41	.95

Reliability Statistics for Listening TW

	Cronbach's Alpha	N of Items
Initial	.88	8
After exclusion of problematic items	.96	6

Item-Total Statistics for Speaking TM

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q52	17.78	28.49	.72	.87
Q53	18.07	29.10	.76	.87
Q54	17.80	29.27	.69	.87
Q55	18.07	30.44	.67	.87
Q56	17.44	31.97	.44	.90
Q57	18.11	31.41	.69	.87
Q59	18.08	30.18	.69	.87
Q62	17.88	30.45	.70	.87

Reliability Statistics for Speaking TM

	Cronbach's Alpha	N of Items
Initial	.89	8
After exclusion of problematic items	.90	7

Item-Total Statistics for Speaking TW

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q58	3.75	1.97	.77	.89
Q60	3.97	1.91	.81	.86
Q61	3.90	1.91	.85	.83

Reliability Statistics for Speaking TW

Cronbach's Alpha	N of Items
.90	3