

Exploring the Shortcomings of the Iranian MSRT English Proficiency Test

Mohammad Reza Ghorbani^{1*}, Hadi Abbassi² and Abu Bakar Mohamed Razali³

¹*Department of Foreign Languages, Faculty of Humanities, University of Bojnord, 9453155111, North Khorasan, Iran*

²*Department of Educational Sciences, Counselling, and Guidance, Faculty of Humanities, University of Bojnord, 9453155111, North Khorasan, Iran*

³*Department of Language & Humanities, Faculty of Educational Studies, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*

ABSTRACT

The Iranian Ministry of Science, Research, and Technology (MSRT) English Proficiency Test (EPT) has been in use since 1992. While the MSRT-EPT is generally claimed to be reliable, valid, and practical, it does not assess speaking and writing skills. In this exploratory study, a qualitative approach was used to examine the MSRT-EPT test-takers experiences and language education experts' beliefs about the test as well as their congruence with each other through semi-structured telephone interviews. Convenience and purposive sampling procedures were used to select 15 participants. Inductive coding method was applied to determine invariant constituents. Then, the constituents were reduced to categories, and finally the categories were clustered into 11 themes. Dependability and validity of the study were established through triangulation, inter-coder agreement, and

member checking technique. The problems associated with the MSRT-EPT and a lack of productive skills included a lack of correspondence between the test content and Ph.D. Candidates' needs, negative washback effect, non-theory-based content, inappropriate listening conditions, and a lack of test items originality. On the other hand, the candidates' and experts' perspectives were highly congruent. In light of these findings, the importance of designing a more comprehensive test including all facets

ARTICLE INFO

Article history:

Received: 16 July 2021

Accepted: 04 October 2021

Published: 30 November 2021

DOI: <https://doi.org/10.47836/pjssh.29.S3.07>

E-mail addresses:

ghorbani@ub.ac.ir (Mohammad Reza Ghorbani)

h.abbasi@ub.ac.ir (Hadi Abbassi)

abmr_bakar@upm.edu.my (Abu Bakar Mohamed Razali)

*Corresponding author

of the language proficiency construct was highlighted, and some suggestions were made for future research.

Keywords: Assessment of Ph.D. candidates, English Proficiency Test (EPT), Ministry of Science, Research, and Technology (MSRT), shortcomings, standardized tests

INTRODUCTION

While Standardized Tests (STs) are playing an increasingly prominent role in higher education decisions in recent years, there has always been a torrent of complaints about them. The criticism and grumbling associated with STs are not new phenomena. Proponents of STs argue that they are fair because they measure student ability objectively. In addition, due to their objectivity, STs can be used for comparison and accountability purposes (Churchill, 2015). However, opponents believe that STs are neither fair nor objective (Singer, 2019; Strauss, 2017) because they cannot measure students' actual progress through a one-time performance evaluation (Martinez & Miller 2018).

The International English Language Testing System (IELTS) and the Test of English as a Foreign Language (TOEFL) are two of the major and widely accepted English proficiency exams for non-native English language speakers intending to enroll in English-speaking universities worldwide. While these two tests differ in format, scoring, approach, and more, they determine students' English proficiency

level by assessing their reading, writing, speaking, and listening skills.

In the Iranian context, English is taught as a foreign language and a subject in high schools and universities. Therefore, Ph.D. students must pass one of the recognized English proficiency tests before graduation. Since taking the TOEFL or IELTS is expensive, the former Ministry of Culture and Higher Education (MCHE) developed a local standardized English Proficiency Test (EPT) known as the MCHE-EPT in 1992. In 2000, the name of the Ministry (MCHE) was changed to the Ministry of Science, Research, and Technology (MSRT). The MSRT was established in 2002. Consequently, the MSRT-EPT is required to be taken by all the Iranian Ph.D. candidates at the state-run universities and higher education institutes, and it is held almost every month. Therefore, this exam is of high importance and has serious consequences for stakeholders.

The MSRT-EPT is a standardized national test to assess the Iranian Ph.D. candidates' overall English as a Foreign Language (EFL) proficiency. This paper-and-pencil test consists listening comprehension, grammar (structure and written expression), and reading comprehension. All three parts of the MSR-EPT consist of multiple-choice questions. The listening comprehension section is comprised of 30 items. Candidates have 30-35 minutes to complete the items. The grammar section is also comprised of 30 items. Candidates have 20 minutes to complete the items. In the reading comprehension section

(four passages usually followed by ten questions), candidates are given 45–50 minutes to answer 40 questions. The total test duration is 1 hour 35–45 minutes. The duration is given in the range depending on the tasks (e.g., the length of the reading comprehension passages), and the allocated time may vary from one test administration to another.

The multiple-choice items in the MSRT-EPT are scored through a computerized scoring system. A test taker's MSRT-EPT score is only valid for two years from the date of taking the test. If the candidates fail to get the required minimum cutoff score (50%), they can register and retake the test without any restrictions. State scholarships are awarded only to candidates who perform above the MSRT-EPT cutoff score (at least 50 out of 100) to continue their studies abroad.

One of the drawbacks of the MSRT-EPT is the probability of guessing the correct answers by test-takers. It is because there are no negative points for wrong answers in the MSRT-EPT. Since test-takers have no marks deducted for giving incorrect answers, this lack of negative points for guessing can lead to chance achievement in test scores (Burton, 2001; Fulcher, 2010). However, Espinosa and Gardeazabal (2010) pointed out that if points were deducted for incorrect answers, test-takers may be cautious and not answer some questions even though they are more likely to choose correct answers.

Noori and Zadeh (2017) state that the MSRT-EPT is generally reliable, valid, and practical. It is well-designed, easily

administered, and objectively scored. The benefits of the test include ease of accessibility, a computerized scoring system, and reasonable fees. However, the test does not assess the productive skills of speaking and writing. It is not based on real-world situations and students' needs. It is administered under different and inappropriate conditions. Since the test is not based on the latest testing trends, many students who pass the test cannot communicate in authentic contexts.

While developing and using tests based on the communicative approach was not possible in the past due to a lack of infrastructure facilities, the communicative assessment of all language skills is readily feasible using information and communication technologies in the 21st century (Yildiz, 2019). Thus, there is a strong need to study the MSRT-EPT shortcomings and help the decision-makers adjust the test to fulfill the requirements of the Iranian context by the emerging trends. Therefore, the following questions are formulated to identify the problems associated with the MSRT-EPT and compare the experts' beliefs with the test-takers experiences:

1. What are the problems associated with the MSRT-EPT based on the language education experts' perspectives and the Ph.D. candidates' experiences?
2. How congruent are the language education experts' perspectives and the Ph.D. candidates' experiences on the MSRT-EPT?

REVIEW OF RELATED LITERATURE

Schmidgall et al. (2019) pointed out that defining the assessment construct (e.g., overall English language proficiency), which is the basis for the meaning of test scores, is one of the key steps in the test development process. However, language proficiency unique to humans in its most complex form is an abstract, invisible ability in the brain, which has nothing to do with how a test is constructed. Language proficiency tests measure how well an individual has mastered a language. There are four domains to language proficiency: reading, writing, speaking and listening.

According to the latest theories, the development of these four integrated skills results from social interaction. Social interaction with the environment plays a key role in cognitive development (Vygotsky, 1978 as in Brown, 2000 & Kaufman, 2004). Larsen-Freeman and Long (1991) pointed out that the interactionist perspectives are better than other theories “because they invoke both innate and environmental factors to explain language learning” (p. 266).

Based on the sociocultural theory (interactionist approach), language emerges from social interaction. According to Bachman (2007), social context and abilities to interact in specific situations form the construct, implying that the construct definition in language assessment inevitably involves presenting ability-in-context. Although Norris (2016) acknowledged that task-based assessment conditions must

approximate real-life contexts to indicate the actual performance of test-takers, the MSRT-EPT lacks the speaking and writing assessment sections. Therefore, one question worth asking is whether the test measures the target construct relating to descriptions of the overall English language proficiency of the Iranian Ph.D. candidates who need to use it to take part in international conferences and publish articles in well-established academic journals. Based on the sociocultural theory, which underpins this study, this test seems to be deprived of the sociocultural features of real-world situations.

According to Purpura (2004), the overall language proficiency conceptualized as a multi-componential ability by many researchers consists of four modalities of listening, speaking, reading, and writing as well as linguistic elements such as vocabulary, grammar, phonology, socio-pragmatics. Powers (2013) asserted that testing English-language skills in all four domains drive teaching and learning and improves the overall communicative competence. Bruce (2018) argued that if an assessment does not adequately measure all facets of the intended phenomenon, construct underrepresentation occurs and detrimentally influences the test use, score interpretation, and evaluation.

Since the MSRT-EPT is a high-stakes test having a profound impact on many stakeholders at the national level, a comprehensive, four-skills assessment is in order. This test is similar to the Iranian National University Entrance Examination (UEE), in which listening,

speaking, and writing skills are not tested. Limiting language assessment to grammar, vocabulary, and reading skills in the UEE has led to a detrimental washback effect on students' English learning activities as well as English teachers' curricular planning and instruction (Ghorbani, 2008; Ghorbani & Neissari, 2015).

A study by Ghorbani et al. (2008) revealed that since test scores in the Iranian educational context provide the only benchmark to assess students' progress in schools, teachers usually rate their students based on their performance in the written exams. They argued that teachers might neglect the oral exams because they tend to teach to the test. The findings of another study by Ghorbani (2012) on the controversy over abolishing the UEE in Iran showed that most informants supported the incremental modification of the UEE. In contrast to the UEE, the MSRT-EPT included the listening section. However, the MSRT-EPT is similar to the UEE, which lacks the speaking and writing sections. Hence, identifying the MSRT-EPT problems is the first step for its modification.

Despite the significance of the MSRT-EPT, only a few studies have been conducted on it. Sahrai and Mamagani (2013) studied the validity and reliability of 10 MSRT-EPTs and found that it generally has acceptable reliability ($p > 0.7$) and validity. However, their study revealed that between the grammar and reading comprehension parts is higher than the correlation between the listening and grammar parts or the listening and reading comprehension parts. Although

the test correlates well with the previously validated and well-established TOEFL, it still requires more substantiation because the TOEFL excludes the speaking skill and measures vocabulary and grammar as separate rather than integrated skills. They believe that the test takers' poor performance in the listening comprehension section of the MSRT-EPT, compared to the reading comprehension and grammar parts, is generally attributed to the listening conditions of the test. They suggested an individual-based listening system to improve the quality of the listening comprehension section of the test. In this study, the informants' perspectives are sought to fill this gap.

Sichani and Tabatabaei (2015) studied the reading comprehension section of the MSRT-EPT using both qualitative and quantitative approaches. The quantitative phase used factor analysis to examine 65 English as a Foreign Language (EFL) students' and 25 experts' perspectives on the reading section. The explanatory factor analysis result did not confirm that the reading section assessed the reading skills. In the qualitative phase, most of the EFL experts and test-takers who were interviewed believed that different items on the reading section of the MSRT-EPT measured the reading ability of the test-takers. While Sichani and Tabatabaei (2015) focused on one section of the test, the present study addresses the test's shortcomings as a whole.

Noori and Zadeh (2017) investigated the strengths and weaknesses of different parts of the MSRT-EPT by analyzing

the test items. They also reviewed the MSRT-EPT related studies conducted to date. They concluded that while the test is generally reliable, valid, well-developed, easily accessible, and less expensive, it still needs more substantiated. To improve the quality of the test, they suggested including the speaking skill, using computerized assessment procedures, considering more integrative communicative items, providing better conditions for testing listening (e.g., using individual-based systems), and penalizing wrong answers (adding a guessing penalty). Each of the suggestions mentioned above is addressed in depth in this study.

Semiyari (2019) studied the MSRT-EPT scores' dependability using G-theory. They examined different sources of variations in isolation (persons, items, sections, gender, and fields of study) and their interactions. The analysis of 1600 pre-intermediate to intermediate participants' performance showed that the test scores were highly reliable. Furthermore, the researchers reported that gender and subject field was negligible, but the difference among persons' performance across items was considerable. This difference probably indicates that high reliability alone is not enough for such an important test.

Each of the studies mentioned above has focused on some specific features of the current MSRT-EPT. Narrowing down a topic and concentrating on its particular aspects can be the strength of a study. While these studies have contributed to a better understanding of the MSRT-EPT, the main

weakness is that they have only addressed what is included in the test. The English language proficiency as a unitary construct, which covers all four language skills, is left under investigation. The current MSRT-EPT, therefore, needs to be examined for its shortcomings.

Based on a critical analysis of the current literature, studies have yet to explore the shortcomings of the MSRT-EPT from the perspectives of experts and test-takers. The current study collected the experts' and test-takers perspectives on the test through in-depth interviews using a qualitative approach and a phenomenological research design. By outlining the rationale for a comprehensive four-domain approach to the target construct assessment, the present study investigated the way the MSRT-EPT is viewed by Iranian language education experts who are aware of the theoretical issues associated with the test. It also investigated former Ph.D. candidates' perspectives as they have experienced the test and are aware of the practical issues.

In sum, this study explores the shortcomings of the present MSRT-EPT in measuring the Ph.D. students' overall communicative competence as a unitary construct. The theoretical perspectives of the experts and the practical perspectives of the test-takers can help testing authorities to improve the quality of the test. Although the studies reviewed show that the test is reliable, they are only limited to reading and listening skills. It is, therefore, necessary to identify the theoretical and practical shortcomings of the MSRT-EPT in terms of

all language skills. Since what is theoretical may be different from what is practical, this study was conducted to identify the problems associated with the MSRT-EPT.

RESEARCH METHOD

Research Design

One possible solution to assess the shortcomings of the MSRT-EPT is to examine the test-takers experiences and experts' beliefs. As emphasized by Edwards and Holland (2013) and Flick (2018), qualitative interviewing is generally used to investigate the experiences and perspectives of the interviewees to gain a better understanding of an issue.

This exploratory study used an interpretive phenomenological and qualitative epistemological approach to address the current MSRT-EPT shortcomings. It investigates the test-takers experiences and language education experts' views regarding the test and their congruence with each other. The phenomenological approach was used to describe the MSRT-EPT test-takers lived experiences, and the exploratory expert interview with an epistemological function (Bogner & Menz, 2009) was used to gain experts' knowledge. In this study, test-takers refer to the Ph.D. students in non-English fields, and experts refer to the English language education lecturers.

Ary et al. (2010) noted that since an experience has different implications for different people, researchers should use phenomenological methods like unstructured interviews to explore the perceptions and

experiences of individuals. Leimeister (2010) believes that epistemology is the basis of appropriate research methods. Epistemology, the study or theory of knowledge, deals with all aspects of knowledge acquisition, including what constitutes knowledge, how knowledge is acquired or produced, and how its transferability can be assessed (Moon & Blackman, 2014). Epistemology was the most suitable approach in this study because it helped the researchers frame their study and discover knowledge.

A combination of the phenomenological approaches (focusing on the study of Ph.D. candidates' lived experiences) and the epistemological approaches (focusing on the discovery of the language education experts' knowledge) in this study helped the researchers address the problems associated with the MSRT-EPT more comprehensively. Furthermore, this research design enabled the researchers to compare and contrast the language education experts' beliefs with the test-takers experiences. Combining these two congruent approaches helped the researchers analyze and triangulate the data from two different sources, thus enhancing the credibility of the research findings and the study's strength.

Sampling and Participants

The convenience and purposive sampling method was used to recruit 15 participants for this study. In this method, since there is no equal opportunity for all qualified individuals in the target population to participate in the study, the study findings

are not necessarily generalizable to the population. The researchers used convenience sampling because the target subjects were nearer and more accessible to them. Purposive sampling was used to select the subjects suited for the study. The participants included eight test-takers (three male and five female) and seven experts (six males and one female). They were the only available subjects that could serve the purpose of the study. Therefore, the selection criteria and justifications for the number of different groups (e.g., male/female) were limited to the subjects' availability and suitability. The study's research objectives determined the choice of participants, and saturation determined the number of participants.

Due to the coronavirus pandemic, the researchers used personal cell phones to approach and recruit the participants. All participants are associated with the University of Bojnord (UB) and the Kosar University of Bojnord, Bojnord, North Khorasan province, Iran. They are academic staff (11 with Ph.D. degrees and four are Ph.D. students). The experts are proficient in English and native-like. They are familiar with the importance, structure, and function of the MSRT-EPT, and their work experience ranges from five to 32 years. The Ph.D. candidates had passed the MSRT-EPT, with at least an intermediate level of English proficiency. Since the researchers did not have permission to use the participants' names, the participants' initials were used throughout the paper.

Data Collection and Analysis

Due to the coronavirus crisis, the data were collected through in-depth telephone interviews during November 2020. The duration of each conversation was about half an hour. In the first phase, eight different semi-structured interviews (eight interviewees were asked the same questions) were held to elicit data from the test-takers about their experiences with the MSRT-EPT. First, the researchers prompted the participants to describe their experiences with the test carefully. Then, after describing the fundamental features of the test-takers common experiences, the researchers were better positioned to explore the experts' perspectives about the test.

In the second phase, seven different semi-structured interviews (seven interviewees were asked the same questions) were conducted to gather data about the test construct (language proficiency), including what it is, how it is acquired, how it is generated, how it is assessed, and when the results are judged to be adequate to claim that it is warranted or justified. There were only two phases of interviews in this study because the first phase addressed the lived experiences of the Ph.D. candidates, and the second phase addressed the theoretical knowledge of the language education experts regarding the MSRT-EPT. Finally, the test-takers and experts' perspectives were assessed to see how congruent they were.

After gathering the related data based on the research questions from the

participants, the researchers transcribed the test-related interviews verbatim. They searched for significant statements that had particular relevance to the MSRT-EPT. The researchers did not use any software for data analysis. The inductive coding method was manually applied to determine the invariant constituents in the data. Then, the constituents were reduced to categories, and finally, the categories were clustered into themes. Data collection and analysis were continued until saturation was reached. That is, further coding was no longer feasible.

The dependability and validity of the study were established through triangulation using two methods—phenomenology and epistemology—to understand the MSRT-EPT shortcomings. They were also enhanced by an inter-coder agreement in which two of the researchers coded the same transcript and compared the results. Furthermore, they were improved by member-checking. That is, the researchers, in the interpretation process, returned the results to the interviewees to review the interpretations and descriptions of the data and check for accuracy.

According to Johnson and Christensen (2017), classical phenomenologists suggest that researchers bracket or suspend their taken-for-granted orientation towards preconceptions about the phenomenon being studied to experience its essence vicariously. The researchers used the Epoche or bracketing technique in this study. In the

phenomenological analysis, it is essential to mitigate the potentially detrimental impact of the researchers' preconceptions that could contaminate the research process. The researchers intentionally set their experiences aside and suspended their own beliefs. They assumed that each interviewee was y unique.

Instead of investigating just the variant part of the data, the researchers sought to understand the essence (commonality or invariant structure) of the experience. The researchers found that certain participants described the MSRT-EPT somewhat differently. While this information was useful in understanding and describing the interesting differences, the researchers were most interested in describing the essence of all the participants. The responses from each participant were considered in the discussion; however, the focus was on the general patterns and findings based on all the subjects' perspectives.

RESULTS

This study benefited from the test-takers experiences and the language education experts' perspectives with regard to the MSRT-EPT problems. The in-depth analysis of the telephone interview results led to some general themes. Based on the test-takers interviews, the following overarching themes were identified for the first research question.

Table 1

The test-takers perspectives on the problems associated with the MSRT-EPT

No	Theme	Frequency
1	Lack of productive skills (speaking and writing)	8
2	Lack of correspondence between the test content and Ph.D. candidates' needs	8
3	No positive washback effect	8
4	Inappropriate listening conditions	5
5	Lack of test items originality	1

Based on the experts' interviews, identified for the first research question. the following overarching themes were

Table 2

The language education experts' perspectives on the problems associated with the MSRT-EPT

No	Theme	Frequency
1	Lack of productive skills (speaking and writing)	7
2	Lack of correspondence between the test content and Ph.D. candidates' needs	7
3	Lack of positive washback effect	7
4	Not based on the latest theories	5
5	Inappropriate listening conditions	5
6	Lack of test items originality	1

The shortcomings of the MSRT-EPT based on the test-takers perceptions and experts' perspectives (Table 1 & Table 2) are combined and summarized in Figure 1 in order of theme frequency to answer the second research question (How congruent are the language education experts' perspectives and the Ph.D. candidates' experiences on the MSRT-EPT?).

As indicated in Table 1 and Table 2, the responses from both participants were in close alignment with each other. The main difference was related to the third theme in Figure 1. While five out of seven experts believed that the test content was not based on the latest theories, the test-takers did not mention this theme.

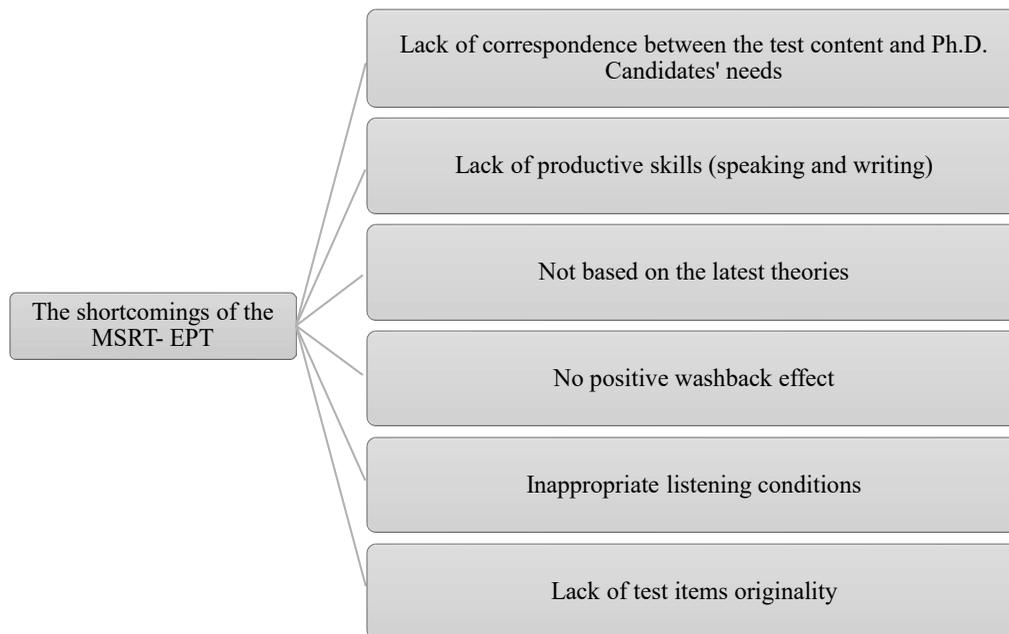


Figure 1. The shortcomings of the MSRT-EPT in order of theme frequency

DISCUSSION

As indicated in Table 1 and Table 2 as well as Figure 1, all the test-takers and language education experts unanimously referred to the problems associated with the MSRT-EPT, including lack of correspondence between the test content and Ph.D. candidates' needs, lack of productive skills, negative washback effect, inappropriate listening conditions, and lack of test items originality. In addition, seven experts referred to the non-theory-based content of the test as the main problem.

A lack of speaking and writing skills suggests that the MSRT-EPT does not adequately measure all aspects of the intended construct (overall language proficiency). Language proficiency has multiple facets. When one of the dimensions

is not used in the measurement, construct underrepresentation occurs and negatively affects the test use, score interpretation, and evaluation (Bruce, 2018). A lack of speaking and writing assessment is the first and most frequent theme. This claim is corroborated by an informant (G. H. K.) as follows: *“Language includes four main skills. Grammar, vocabulary, and pronunciation are subskills. Listening and reading alone cannot assess language proficiency. Test-takers are not required to produce anything in the MSRT-EPT. There is no speaking and writing section”*.

This finding is inconsistent with the results of a previous study by Sahrai and Mamagani (2013), who studied the validity and reliability of 10 MSRT-EPTs and claimed that the test generally

possesses acceptable reliability and validity. It also contradicts Noori and Zadeh's (2017) conclusion that the MSRT-EPT is generally reliable, valid, and well-developed. However, it is consistent with their suggestion that including the speaking and writing components can improve the quality of the test as a whole. Therefore, although the test has acceptable reliability and validity, including the productive skills can improve these two key characteristics.

The second most frequent theme is that a lack of correspondence between the test content and Ph.D. candidates' needs. It is confirmed by a respondent (G. H. K.) who mentioned that: *"Ph.D. students are highly expected to write articles in English and take part in international conferences. The test should be comprehensive and based on the candidates' needs. These reading and listening multiple-choice items are not based on the real-world context. Writing and speaking are very important for Ph.D. students. I think, at least, writing should be added to the content of the test"*. It implies that if the test were designed based on the candidates' needs, it would probably reinforce the test-takers motivation and encouragement.

This finding aligns with Powers' (2013) argument that communicative competence is key in English-language proficiency and involves all main language skills (reading, listening, speaking, and writing). In addition, Ph.D. candidates need to use English for article writing and take part in international conferences. Therefore, the four language domains should be included

in the test in an integrative way to satisfy all the needs of the intended test-takers.

A lack of positive washback effect is the third most frequent theme emphasized by all participants. One of the experts (M. E. S.) elaborated on this problem: *"The test-takers just focus on mastering decontextualized grammar and vocabulary. They try to attend private language institutes and classes where they just study grammar and vocabulary books like 504 and 1100 Words so that they can boost their test-taking strategies based on what appears on the test. They cram for the exam. They just want to pass the exam and meet the Ph.D. requirements."*

It is in line with previous research findings regarding the negative washback effect of high-stakes tests on both teachers and students in the Iranian context (Ghorbani, 2008; Ghorbani et al., 2008; Ghorbani & Neissari, 2015). However, the MSRT-EPT washback effect is different from that of other nationwide Iranian tests like the UEE, where teachers adjust their teaching methods, and learners adjust their learning activities based on what appears in the test. University professors do not teach any specific materials related to the MSRT-EPT to be affected by the test format. However, future Ph.D. candidates will focus on the features that appear in the test and ignore what is excluded. Hence, if the test is designed to exert a positive effect, it can be used as a driving force to promote useful learning activities.

The fourth important problem identified from the responses was inappropriate

listening conditions during the test. Ten out of 15 participants referred to this challenge. One of the test-takers (S. K.) stated that: *“Loudspeakers did not function very well. It took the test-administrators a few minutes to adjust and operate them. The voice was not clear. It was vague. It was interrupted. Test-takers were stressful. Loudspeakers were too close to test-takers. A few candidates were next to the loudspeakers. They were not able to understand anything. It was not a good environment for listening at all.”*

A language expert (M. E. S.) who had taken the MSRT-EPT a few years ago also reported that: *“The test-takers were not satisfied with the listening conditions at all. I took this test in 2009. I really could not understand what I heard. English Language students have to get 80 out of 100 to meet the Ph.D. requirements. It is too difficult to get the pass mark even for the English language candidates because the quality of the listening part is not acceptable.”* When the testing conditions are different for different candidates, the reliability and validity of the test are questionable.

This finding is in line with Sahrai and Mamagani's (2013) recommendation as well as Noori and Zadeh's (2017) suggestion that providing better conditions for testing listening skills through using individual-based systems is essential to improve the quality of the test. Educational technology advancements increasingly make it more feasible to use computer-based testing with individual headsets to provide better listening conditions. As the IELTS is held

under the same condition for all candidates in Iran, the MSRT-EPT can also be held under the same condition.

Five out of seven language education experts believed that the test was not based on the latest theories. None of the test-takers referred to this theme because they were probably unaware of the assessment theories and just reflected on their experiences. One of the experts (J. Z.) stated that: *“Items in a test should not be isolated. Language skills are integrated in the real world. In the MSRT-EPT, it is quite clear that skills are not tested in an integrative manner. For example, grammar cannot be separated from reading and writing. It is better to test grammar in students' writing and reading.”*

This finding is consistent with the sociocultural theory that language emerges from social interaction. Social context and abilities to interact in specific situations form the construct. That is, construct definition in language assessment inevitably involves presenting ability-in-context (Bachman, 2007). The test in the present format is not based on real-world situations in which there is a natural interaction and meaningful communication.

A lack of originality in the test items is the last theme. Only one test-taker and one expert referred to this problem. One of the experts (M. A. R.) mentioned that: *“All test items have already been used elsewhere. There is no board of exam to develop original items. The test is not standardized. The items are probably taken from the TOEFL and IELTS samples.”* It implies

that the test-takers might pass the test by studying sample tests in which some items are sometimes included in the real test.

Furthermore, a test-taker (H. D.) stated, *“I think you can find all the listening section in the market. The items are taken from the TOEFL. Recently it seems that they have tried to change the test and improve it. However, I think test-takers will find the reference of the test items. I think developing a language proficiency test at the national level is too difficult. The test developers have to resort to international standardized test samples to design their local tests.”* These findings indicate that since the test items are not original, some test-takers may just review previous sample tests and manage to get a pass. It will, in turn, affect the validity of the test. Furthermore, the listening section is difficult to be developed by non-native English speakers. Due to this, they use the available listening material developed by native speakers.

To answer the second research question (How congruent are the language education experts' perspectives and the Ph.D. candidates' experiences on the MSRT-EPT?), the test-takers experiences and experts' perspectives were compared and contrasted. As indicated in Tables 1 and 2, the responses from both participants were in close alignment with each other. The main difference was related to a theme that five experts had emphasized. While five out of seven experts believed that the test content was not based on the latest theories, none of the eight test-takers referred to any theories.

Since the test-takers field of study was not related to second language education, they might be unaware of the assessment theories. However, they shared test-related problems without mentioning any theories, which indicate a lack of congruence between the test content and the latest theories. For example, the following quotation from one of the test-takers (M. S. H.) confirmed this claim: *“This test lacks speaking. It does not help us improve our communication abilities. It is of no use for enhancing letter or article writing. It is for improving translation only. The test must include a writing section with open-ended items.”* This statement refers to the fact that the test's content does not represent the construct of interest accurately.

While the experts' beliefs focused on the theoretical aspects of the MSRT-EPT, the test-takers experiences focused on the practical problems of the test in this study. The MSRT-EPT is mandatory for all Ph.D. students regardless of their field of study; however, the passing mark is at least a score of 80 out of 100 for English major students and 50 out of 100 for the rest of Ph.D. candidates. Therefore, it may justify why the experts' perspectives are somehow similar to the Ph.D. candidates' perceptions.

In sum, the findings of this study suggest that the MSRT-EPT in the present form underrepresents the construct of language proficiency. Furthermore, a lack of alignment between the test content, objectives, the latest related theories, and Ph.D. candidates' needs has led to a detrimental washback effect. Therefore, it

is high time for language education policy-makers and testing authorities to revise and improve the quality of the test based on the latest research findings.

CONCLUSION

This exploratory study examined the MSRT-EPT test-takers experiences with the test and the language education experts' beliefs about the test's shortcomings through semi-structured telephone interviews. The findings revealed that the test is not designed in alignment with the latest theories or the findings of recent studies on language education and assessment. Moreover, since it does not assess productive speaking and writing skills, it underrepresents the construct it claims to measure. A lack of alignment between the test content and its objectives has a pernicious washback impact on the Ph.D. candidates and the related instructors. Inappropriate listening conditions in some testing centers are another serious problem, which has made the test more biased.

These findings highlight significant implications for foreign language policy-makers, testing authorities, test developers, and test-takers. The evaluation based on the test outcome without considering the internal and external factors that affect the reliability and validity of the test may result in incorrect interpretation and decisions. Consequently, as pointed out by most participants, identifying the problems and inadequacies associated with the current test and accordingly rectifying them can improve the validity of score interpretations.

When the quality of the MSRT-EPT is improved, the decision-making will be enhanced accordingly.

In sum, this study highlighted the importance of designing a more comprehensive test, including all facets of the language proficiency construct. The findings of the study contributed to the betterment of the MSRT-EPT in the future. Despite the identified shortcomings, the test can be redesigned and improved to include the productive skills of speaking and writing. Although these skills were difficult to assess in an integrative way at the national level in the past, the development of educational technologies has made it possible to easily include such skills in the test to cater to the needs of the intended candidates.

Testing authorities can improve the quality of the listening conditions, which are different from one context to the other, by using computer-based testing. Each candidate is provided with their headsets. In addition, acoustic standards can be used to create a good listening environment. Poor acoustics in some test settings makes it difficult for test-takers to make the best use of their listening time. Ongoing refurbishment is a chance to modify and improve the acoustic conditions of testing centers.

The current study was limited to the views of eight test-takers and seven language education experts on the problems associated with the MSRT-EPT. The study served as a preliminary investigation of the test. Further quantitative and qualitative studies involving a larger and wider group

of stakeholders are suggested to support or reject the perspectives demonstrated in this study and uncover other test dimensions. It is hoped that future studies provide more insights for the improvement of this high-stakes test. In addition, some of the research topics not covered in this study or ignored in previous studies need further investigation. They are as follows:

To date, no study has addressed the extent to which the test-takers success in the MSRT-EPT can enable them to write scientific papers in English or take part in international conferences. Few studies have dealt with the reasons for not testing in all four domains. The extent to which the testing conditions can affect the MSRT-EPT candidates' performance needs to be investigated. As English language testing authorities have relied on the MSRT-EPT results for decision-making for years, the extent to which such decisions are sound must be investigated.

Since a lack of productive skills in the MSRT-EPT exerts a negative washback effect, future studies can address the degree of the MSRT-EPT washback impact. In addition, future studies can focus on how a comprehensive approach to testing English language proficiency can be implemented. Finally, researchers can study whether it is justifiable to use the MSRT-EPT instead of the well-established standardized tests like the TOEFL and IELTS in the Iranian context.

ACKNOWLEDGEMENT

The authors would like to thank all the participants for their valuable time. They would also like to express their gratitude to the respected reviewers and editors for their appropriate and constructive suggestions.

REFERENCES

- Ary, D., Jacobs, L. C., & Sorensen, C. (2010). *Introduction to research in education*, (8th ed.). Cengage Learning.
- Bachman, L. F. (2007). What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 41-72). University of Ottawa Press. <https://doi.org/10.2307/j.ctt1ckpcf.9>
- Bogner, A., & Menz, W. (2009). The theory-generating expert interview: Epistemological interest, forms of knowledge, interaction. In A. Bogner, B. Littig, & W. Menz (Eds.), *Interviewing experts* (pp. 43-80). Palgrave Macmillan. https://doi.org/10.1057/9780230244276_3
- Brown, H. D. (2000). *Principles of language learning and teaching*. Longman.
- Bruce, B. F. (Ed.). (2018). *The SAGE encyclopedia of educational research, measurement, and evaluation*. SAGE Publications. <http://dx.doi.org/10.4135/9781506326139.n144>
- Burton, R. F. (2001). Quantifying the effects of chance in multiple choice and true/false tests: Question selection and guessing of answers. *Assessment and Evaluation in Higher Education*, 26(1), 41-50. <https://doi.org/10.1080/02602930020022273>
- Churchill, A. (2015). *Bless the tests: Three reasons for standardized testing*. The Thomas B. Fordham Institute. <https://fordhaminstitute.org/>

- national/commentary/bless-tests-three-reasons-standardized-testing
- Edwards, R., & Holland, J. (2013). *What is qualitative interviewing?* Bloomsbury Publishing. <https://doi.org/10.5040/9781472545244>
- Espinosa, M. P., & Gardeazabal, J. (2010). Optimal correction for guessing in multiple-choice tests. *Journal of Mathematical Psychology*, 54(5), 415-425. <https://doi.org/10.1016/j.jmp.2010.06.001>
- Flick, U. (2018). Doing qualitative data collection - Charting the routes. In U. Flick (Ed.), *The SAGE handbook of qualitative data collection* (pp. 3-16). SAGE Publications. <https://doi.org/10.4135/9781526416070.n1>
- Fulcher, G. (2010). *Practical language testing*. Hodder Education/Routledge.
- Ghorbani, M. R. (2008). The washback effect of the university entrance examination on Iranian English teachers' curricular planning and instruction. *The Iranian EFL Journal*, 2, 60-87.
- Ghorbani, M. R. (2012). Controversy over abolishing Iranian university entrance examination. *Asian Education and Development Studies*, 1(2), 139-152. <https://doi.org/10.1108/20463161211240115>
- Ghorbani, M. R., & Neissari, M. (2015). Washback effect of the Iranian Concours on senior high school students' EFL learning activities. *Iranian Journal of Language Testing*, 5(1), 1-28.
- Ghorbani, M. R., Arshad, A. S., Sahandri, M., & Nooreen, N. (2008). The washback impact of the Iranian university entrance examination on pre-university English teachers. *The Journal of Asia TEFL*, 5(3), 1-29.
- Johnson, R. B., & Christensen, L. B. (2017). *Educational research: Quantitative, qualitative, and mixed approaches* (6th ed.). SAGE Publications.
- Kaufman, D. (2004). Constructivist issues in language learning and teaching. *Annual Review of Applied Linguistics*, 24, 303-319. <https://doi.org/10.1017/S0267190504000121>
- Larsen-Freeman, D., & Long, M. (1991). *An introduction to second language acquisition research*. Longman.
- Leimeister, S. (2010). Epistemological foundation and research methodology. In *IT Outsourcing Governance* (pp. 9-13). Gabler Verlag. https://doi.org/10.1007/978-3-8349-6303-1_2.
- Martinez, S., & Miller, R. (2018, November 26). Measuring growth in test scores is key to understanding student progress. *EdSource*. <https://edsources.org/2018/measuring-growth-in-test-scores-is-key-to-understanding-student-progress/604959>
- Moon, K., & Blackman, D. (2014). A guide to understanding social science research for natural scientists. *Conservation Biology*, 28, 1167-1177. <https://doi.org/10.1111/cobi.12326>
- Noori, M., & Zadeh, S. S. H. (2017). The English proficiency test of the Iranian ministry of science, research, and technology: A review. *International Journal of English Language & Translation Studies*, 5(3), 21-26.
- Norris, J. (2016). Current uses for task-based language assessment. *Annual Review of Applied Linguistics*, 36, 230-244. <https://doi.org/10.1017/S0267190516000027>
- Powers, D. E. (2013). Assessing English language proficiency in all four language domains: Is it really necessary? In D. E. Powers (Ed.), *The research foundation for the TOEIC tests: A compendium of studies* (pp. 1.1-1.7). Educational Testing Service.
- Purpura, J. (2004). *Assessing grammar*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511733086>

- Semiyari, S. R. (2019). Investigating MSRT (MCHE) scores' dependability by means of G-theory: Variations due to persons, items, sections, gender, and subject fields. *International Journal of Language Testing*, 9(2), 56-81.
- Sahrai, R., & Mamagani, H. (2013). The assessment of the reliability and validity of the MSRT proficiency test. *The Educational Assessment Journal*, 10(3), 1-19.
- Schmidgall, J., Oliveri, M. E., Duke, T., & Grissom, E. C. (2019). *Justifying the construct definition for a new language proficiency assessment: The redesigned TOEIC Bridge® tests - Framework paper* (Research Report No. RR-19-30). Educational Testing Service. <https://doi.org/10.1002/ets2.12267>
- Sichani, E. F., & Tabatabaei, O. (2015). Construct validity of MSRT reading comprehension module in Iranian context. *English Language Teaching*, 8(9), 173-186. <https://doi.org/10.5539/elt.v8n9p173>
- Singer, S. (2019, July 5). *Gadfly on the Wall Blog: Standardized tests are not objective measures of anything*. National Education Policy Center (NEPC). <https://nepc.colorado.edu/blog/standardized>
- Strauss, V. (2017, April 19). 34 problems with standardized tests. *The Washington Post*. <https://www.washingtonpost.com/news/answer-sheet/wp/2017/04/19/34-problems-with-standardized-tests/>
- Yildiz, T. E. T. (2019). Emergent trends and research topics in language testing and assessment. *Akdeniz Eğitim Araştırmaları Dergisi*, 13(28), 66-81. <https://doi.org/10.29329/mjer.2019.202.4>