

Test Administration Conditions of the General English Section of the Iranian National PhD Entrance Exam: Are the PhD Exam Candidates Satisfied?

Shiela Kheirzadeh¹, S. Susan Marandi², Mansoor Tavakoli³

Received: 25 July 2015

Accepted: 01 September 2015

Abstract

Test administration conditions, namely, the timing of the test, the testing venues and the exam proctors/inspectors are influential factors that may introduce construct-irrelevant variance to a test, if ignored, and therefore render a test invalid, especially in the case of high-stakes tests. The purpose of the present study was to investigate the satisfaction of the PhD exam candidates with the administration conditions of the General English section of the National Iranian PhD Entrance exam, focusing on the three above mentioned factors. The timing factor was investigated from the perspectives of suitability of the time, that is, morning administration, evening administration and time delay in administration. The testing venue included the location of testing venues, commuting, finding seats, ventilation and lighting of the venues and finally, the proctor/inspector factor looked into the presence of the inspectors, satisfaction with the behavior of the proctors, peacefulness of the session, refreshments and finally the possibility of cheating. The data were collected by a sixteen item questionnaire which was distributed among 173 PhD exam candidates, both males and females, in 30 different testing venues all over Iran. The results indicated that they were satisfied with all three factors and the entire exam administration process; however, they made some comments to improve the exam administration process.

Key words: *Test Administration, Testing Venue, Timing, Proctors/Inspectors, Iranian National PhD Entrance Exam*

¹English Department, Alzahra University, Tehran, Iran (Corresponding Author). Email: sh.kheirzadeh@yahoo.com

²English Department, Alzahra University, Tehran, Iran. Email: susanmarandi@alzahra.ac.ir

³English Department, University of Isfahan, Isfahan, Iran. Email: mr.tavakoli14@gmail.com

as standardized conditions (Wise et al., 2010). According to Xi (2010) and Kunnan (2000, 2003), among the factors that may create construct-irrelevant variances and render the test invalid are inconsistent test administration, lack of accommodation for test takers with disabilities, and raters' bias. Therefore, if we want the results of our tests to be valid and dependable, the test needs to be of a high quality both in development and administration (Saville, 2012).

Since the 1980s, researchers and scholars have become interested in developing high-quality tests for practical purposes; as a result, a large number of textbooks (e.g. Bachman & Palmer 1996, 2010; Bachman 1990; Weir 2005; Alderson et al., 1995; Hughes, 2003) were written to guide the test practitioners through the daunting task of test development and administration. A common feature among all these textbooks, according to Saville (2012), is that test development is conceived of as a process including a series of logical steps. These steps begin with *test design* and *item writing* and then move through the *administration* and *reporting* the results. The final stage of this assessment cycle is the *evaluation* step. Moreover, in a different classification, Bachman and Palmer (1996) defined a three-stage model including *design*, *operationalization* and *administration*. However, what has been concluded by Saville (2012, p. 402) as an overall assessment cycle is a cycle that has the following five main stages:

- Planning and design following a decision to develop a test;
- Development, including initial development and systems for test assembly;
- Delivery, including routine test assembly and the administration of the assessment;
- Processing, including marking and grading and issue of results;
- Review and evaluation.

As it is evident in the abovementioned classifications, the *administration* step is considered as a main stage in the test development process. When test administration is discussed in the testing related texts, the only concern is delivery system (Fulcher & Davidson, 2007) which is often related to test taking processes or the physical environment (Alderson et al., 1995; Bachman 1990; Bachman & Palmer, 1996). However, according to Saville (2012), just in rare instances the issue of test administration is dealt with when the "operational management" of the tests has been considered. Of these few cases, Fulcher and Davidson (2007, p. 115), in a chapter devoted to test administration stated that, "the delivery of tests requires an infrastructure, and a set of procedures that are usually followed". They further continue that, "for language testing, part of the validity claim is that the test administration and all the processes used by the testing agency are done according to standardized procedures" (p. 127).

2. Literature Review

Standardization is a crucial matter in test administration. Therefore, it must be guaranteed if different venues, times or groups of people are involved. Moreover, fair tests are those that are both unbiased and meet recognized ethical and administration standards. For example, standardized tests should be administered exactly according to their application instructions and everyone should receive the same instructions (Zieky, 2006). Along the same lines, it has been stated in *the Standards for Educational and Psychological Testing* (AERA 1999, p. 61) that:

The usefulness and interpretability of test scores require that a test be administered and scored according to the developer's instructions. When directions to examinees, testing conditions, and scoring procedures follow the same detailed procedures, the test is said to be standardized. Without such standardization, the accuracy and comparability of score interpretations would be reduced. For tests designed to assess the examinee's knowledge, skills, or abilities, standardization helps to ensure that all examinees have the same opportunity to demonstrate their competencies. Maintaining test security also helps to ensure that no one has an unfair advantage.

In administering the tests on a large scale, the personnel in the testing venues are in charge of the test administration. Accordingly, the following points, as stated by Saville (2012, p. 406), should be clearly described and specified.

- *The physical setting* – how the venue is selected and how the rooms where the assessments take place are managed, including the safety and security of the premises;
- *The type and number of personnel*– needed to oversee the administration on the day of the assessment to ensure safety and guarantee test security;
- *The management of the interaction between test takers and administrators before, during and after the assessment takes place* – including the arrival at the venue of the test taker, the checking of their identities and eligibility, the seating arrangements, provision of accommodations to meet special needs or requirements;
- *The management of the assessment procedures themselves* – providing instructions, handing out and collecting test papers and answer sheets, monitoring for malpractice (including all kinds of cheating), ensuring that timings are respected.

As stated by Fulcher and Davidson (2007), the personnel or staff needs to have training in how to deal with the test takers, including checking their identity, giving instructions, managing the late arrivals, disruptive behaviors, those who finish early and especially cheaters. Considering cheating, Cizek (2003) identifies three types of cheating: (a) cheating by giving, taking, or receiving information from others; (b) cheating through use of prohibited materials; and (c) cheating by impeding the testing process. These types of cheating are observed in

traditional paper and pencil testing as well as online testing. Studies by Crown and Spiller (1998), McCabe et al. (2001) and McInerney (2008), to name a few, confirmed the view that a very high inclination toward cheating has been found among the test takers in the last decades. As a result, *Standards for Educational and Psychological Testing* (AERA 1999, p. 64) asserted that, “reasonable efforts should be made to ensure the integrity of test scores by eliminating opportunities for test takers to attain scores by fraudulent means”.

Moreover, the distribution of the collection of test materials should be timely and standardized. However, since any deficiencies in the administrative procedure may affect the validity and the quality of the tests, it is recommended for the inspectors and visitors from the test development organization to inspect and monitor the personnel and proctors (Saville, 2012). From the above mentioned points, it can be understood that personnel is a major factor in a standardized test administration.

Another critical issue in test administration is the testing venue or testing environment. *The Standards for Educational and Psychological Testing* (AERA 1999, p. 63) stated that, “the testing environment should furnish reasonable comfort with minimal distractions”. Shulman et al. (2011) found that venue and the measurement format affect the performance of test takers on a political knowledge test. Furthermore, Cohen and Wollack (2006, p. 357) mentioned that:

It is reasonable to expect that administration conditions have a non-eligible effect on examinee performance. If it is too hot or too cold in the testing room, performances of some examinees are likely to be negatively affected. Similarly, if it is too noisy, some examinees may be distracted and perform below their potential.

Few studies have focused on the effect of time on standard test administration (Mandinach et al., 2005). Considering the morning or evening administration, Monk (1990) recommended administering adult tests in the morning. The reason is that the temperature peak for energy is morning; therefore, the optimal performance time in a day is morning for adults (Anderson et al., 1991; Webb, 1982). Leigh and Reynolds (1982) studied the effect of time of the day on intelligence test performance. The results indicated no significant difference in either performing the test in the morning or evening except for the verbal section which favored a morning administration. In addition, the results of Gupta’s (1990) study on administering a test among college students indicated no difference in the performance considering the time of the day; however, more research need to be carried out considering time, especially in the case of the high-stakes proficiency tests.

Concluding the abovementioned studies and standards, a standardized test administration process can be viewed from the three perspectives of time, testing venue and the proctors/inspectors. As the title suggests, the present study concerns the administration of the General English section of the National PhD Entrance Exam in Iran. The reason for investigating

the administration process of this exam is that it is a high-stakes and nationwide exam in Iran and the acceptance or non-acceptance of post-graduate students at the PhD level in all academic disciplines is decided by the PhD Entrance Exam; hence, its administration should be taken into account to avoid the introduction of any construct-irrelevant variance that may affect test scores and ultimately the high-stakes decisions made. It is worth mentioning that this is a pioneering study in the context of Iran and is also among the few studies devoted to the investigation of the test administration conditions from the perspectives of test takers. To sum up, the present study is an attempt to find the answer to the following questions.

Q1. Are the PhD exam candidates satisfied with the timing of the exam? Is there any difference between male and female candidates in this regard?

Q2. Are the PhD exam candidates satisfied with the testing venues of the exam? Is there any difference between male and female candidates in this regard?

Q3. Are the PhD exam candidates satisfied with performance of the proctors/inspectors of the exam? Is there any difference between male and female candidates in this regard?

Q4. Are the PhD exam candidates totally satisfied with the test administration? Do they have any comments to improve test administration?

3. Methodology

3.1 Participants

The participants of the present study (n=173) were both males (n=100) and females (n=73) from among the National PhD Entrance Exam candidates in Iran. Their age ranged from 25 to 50. The participants were randomly selected from 30 testing venues in Iran. The reason for having both males and females in the study is that the testing venues of female participants are separate from those of the males; therefore, both groups were considered and compared.

3.2 Instrumentation

A sixteen-item multiple choice questionnaire was developed to measure the three most crucial test administration concerns that may impact the test performance and results. These three factors were time, proctors/inspectors and the testing venues, which were the basis of the first three research questions of the present study. The last research question intended to explore test taker's overall dissatisfaction with the test considering the above mentioned factors, as well as their possible recommendations for the betterment of the exam.

Of these sixteen items, four items dealt with the issue of time, five items investigated the degree of satisfaction of the test takers with the testing venues, and five items concerned the proctors/inspectors. The last two items in the questionnaire explored the overall dis/satisfaction with the whole process of test administration, and requested the test takers to add their suggestions and recommendations, if any, for the better administration of the exam. One point

that needs to be mentioned is that the questionnaire was prepared in the PhD exam candidates' first language, which is Persian. The reliability of the questionnaire, estimated by Cronbach's Alpha, was .61. The validity of the questionnaire was confirmed by expert judges.

3.3 Procedure

The questionnaire was distributed, either manually or through email, among the PhD exam candidates. In both cases, the participants were totally informed about the purpose and the manner of answering. There was no time limitation for filling in the questionnaire.

4. Results

As mentioned above, the first research question of the study attempted to discover if the PhD exam candidates were satisfied with the time of the exam and whether male and female exam candidates differ in their degree of satisfaction. Four questions were devoted to the issue of timing, the responses to which are presented in Tables 1 and 2, below.

Table 1. Satisfaction with timing of the PhD entrance exam

Item	Yes(%)	No(%)	Undecided(%)
Was the administration time suitable?	79	15.6	4.9
Do you prefer the exam to be administered in the morning?	72	24.9	3.1
Do you prefer the exam to be administered in the evening?	24.9	72	3.1
Was there any time delay in the administration?	34.7	60.1	5.2

As can be seen from Table 1, the participants answered four questions concerning the timing issue. The first question was to know if the administration time was suitable. Of the total 173 PhD Entrance Exam candidates, both males and females, 79% were satisfied with the time of the test, the test was administered as 8:00 a.m. However, 15.6% were dissatisfied with this time and 4.9% had no idea.

The next question relating to the timing issue was the preference for the morning administration of the exam. 72% preferred to take the exam in the morning; however, 24.9% did not prefer it and 3.1% had no idea about the suitability of morning administration.

The third timing question was to check if the PhD Entrance Exam candidates prefer the exam to be administered in the evening. The majority of the exam candidates, 72%, disagreed with the idea of the evening test administration and only 24.9% opted for the evening administration. As above, 3.1% have no special comment or idea to share.

The last timing issue question was to find if there was any time delay in the administration of the exam. Majority of the PhD Entrance exam candidates mentioned not time delay, 60.1%, while 34.7% experienced the time delay in the administration process. 5.2% had no idea whether there was a time delay in the administration.

Table 2. Male and female candidates' satisfaction with timing

Item	Male			Female		
	Yes(%)	No(%)	Undecided(%)	Yes(%)	No(%)	Undecided(%)
Was the administration time suitable?	43.3	11.54	4.3	35.55	4.04	0.6
Do you prefer the exam to be administered in the morning?	42.08	12.5	1.45	29.95	12.5	1.47
Do you prefer the exam to be administered in the evening?	12.5	42.08	1.5	12.5	29.32	2.15
Was there any time delay in the administration?	21.4	32.9	3.5	13.3	27.2	1.7

As it was mentioned in the participants' section, males and females were compared with each other, since male and female PhD Exam candidates took the exam in separate testing venues. Therefore, each of the timing questions was separately analyzed for each gender to check if there was any difference in their satisfaction with the timing issue (Table 2).

Considering the suitability of administration time, 43.3% of male PhD exam candidates considered the time suitable comparing with 35.55% for females. Those who perceived the time unsuitable were 11.54% of males and 4.04% of females. To find if there is any significant difference between males and females in their idea about the suitability of the administration time, the Fisher's Exact Test was run. The result of .085 for the Exact Sig. (2-sided) indicated no significant difference between the two genders.

As it was stated in Table 1, 72% of the exam candidates preferred the morning administration of the exam of which 42.08% was for males and 29.95% for females. However, 12.5% of both genders did not favor morning administration. The result of the Fisher's Test (Sig= .86) indicated no significant difference between genders concerning their ideas about morning administration.

As to the evening administration, of the total of 60.1% who disagreed with the evening administration, 42.08% were males and 29.32% were females. However, 12.5% of both genders favored the evening administration. No significant difference was observed between both genders on their idea about the evening administration (Sig = .79).

Thirty two percent of males did not experience the time delay in the administration of the exam in their testing venues while this figure was 27.2% for females. 13.3% of females and 21.4% of males experienced the time delay in administration. However, their ideas did not differ significantly (Sig = .412).

The second research question was concerned with the satisfaction degree of the PhD Exam candidates with the testing venues and if there was any difference between males and females in this regard. Five questions (Tables 3 and 4) were inquiring about the testing venues.

Table 3.Satisfaction with the testing venues of the PhD entrance exam

Item	Yes(%)	No(%)	Undecided(%)
Were you appropriately informed about the testing venues' location?	90.8	6.4	2.3
Could you commute easily to the testing venues?	23.7	75.1	1.2
Could you find your seat easily?	90.8	7.5	1.7
Were the testing venues well-ventilated?	68.8	23.7	7.5
Was there enough light in the testing venues?	86.1	11.6	2.3

The first question among the five questions concerning the testing venues was asking if the PhD exam candidates were appropriately informed about the location of the testing venues (Table 3). Of the total of 173 participants, the majority, 90.8% stated that they had enough information about the testing venue location. 6.4% were not well-informed and 2.3 were undecided.

75.1% of the participants could not easily commute to the testing venues; 23.7% has no problem commuting and 1.2% had no idea. The majority of the participants, 90.8% could find

their seats easily in their testing locations while 7.5% had trouble finding their seats and 1.7% were undecided.

Ventilation of the testing venues was satisfactory to 68.8% of the participants while 23.7% were dissatisfied with it. 7.5% of the exam candidates had no idea to share. Lighting was adequate from the perspective of 86.1% of the participants while 11.6% were dissatisfied with it. 2.3% of the participants had no idea.

Table 4. Male and female PhD entrance exam candidates' satisfaction with the testing venues

Item	Male			Female		
	Yes(%)	No(%)	Undecided(%)	Yes(%)	No(%)	Undecided(%)
Were you appropriately informed about the testing venue location?	51.4	4	1.7	39.3	2.3	.6
Could you commute easily to the testing venue?	12.1	44.5	1.2	11.5	30.5	.2
Could you find your seat easily?	51.4	5.8	.6	39.3	1.7	1.2
Was the testing venue well-ventilated?	37	16.2	4.6	31.8	7.5	2.9
Was there enough light in the testing venue?	49.1	8.1	6	37	3.5	1.7

The male PhD exam candidates were somewhat more satisfied (51.4%) than females (39.3%) with the manner of providing information on the testing venue location; however, the difference between the two genders was not statistically significant considering the Fisher's Exact test Sig of .760. Commuting to the testing venues was more difficult for males (44.5%) than females (30.5%), yet the difference was not significant (Sig = .372). Furthermore, male had less problems finding their seats (51.4%) comparing with females (39.3%) though the difference between them was not statistically significant (Sig = .242).

Males were more satisfied (37%) with the ventilation of their testing venues than female PhD exam candidates (31.8%). Of the total male participants, 16.2% were not satisfied with the ventilation and this percent was 7.5 for females. As it is evident in the percentages, the difference, confirmed by the Fisher's Exact Test sig of .142, was not statistically significant between males and females. The last question concerning the testing venues was checking if the lighting was adequate at the time of administration. 49.1% of "yes" answer for males in comparison with 37% for females indicates that males were more satisfied with the lighting of their testing venues. However, 8.1% of the male population, in comparison with 3.5% of female population, was dissatisfied with the lighting.

The third research question had to do with the proctors and inspectors of the PhD Entrance Exam. Of the total sixteen questions of the interview, five were concerned with the proctors and their responsibilities (Tables 5 and 6). Similar to the analysis of the previous research questions, Table 5 presents the satisfaction percentage of the participants and Table 6 focuses on each gender in detail.

Table 5. Satisfaction with the proctors/inspectors of the PhD entrance exam

Item	Yes(%)	No(%)	Undecided(%)
Were the inspectors available at the exam session?	72.8	22.2	5
Did the exam proctors behave well?	85	4.6	10.4
Could the proctors keep the exam session peaceful?	79.2	13.3	7.5
Did the exam proctors provide the refreshments?	39.9	50.9	9.2
Was cheating possible at the exam?	22.5	66.5	11.0

The first question among the five questions dealing with the issue of proctors/inspectors was to inquire if the inspectors from the exam administration organization, the National Organization for Educational Measurement of the Islamic Republic of Iran, were present at the exam. Referring to Table 5, 72.8% of the participants claimed that the inspectors were present while 22.2% claimed that there was no inspector in their testing venues; moreover, 5% of the exam candidates had no idea.

The second question was asking about the way proctors behaved toward the exam candidates. Eighty five percent of the 173 participants of the study who were from 30 different testing venues throughout Iran were satisfied with the behavior of the exam proctors; however, 4.6% were dissatisfied and 10.4% had no idea in this regard.

79.2% of the participants of the present study stated that the proctors could keep the exam session peaceful enough while 13.3% were displeased with the performance of the proctors in keeping the exam session quiet and 7.5% had no idea.

Satisfaction with the refreshments was probed in the fourth question. Half of the participants of the study, 50.9%, were not satisfied with refreshments and the proctor/personnel performance in this regard while 39.9% were pleased with it. 9.2% had no idea to share. The last question in the series of the proctors/inspectors questions was questioning about the possibility of cheating at the exam session. 66.5% mentioned that cheating was not possible in their testing venues due to the satisfactory performance of the exam proctors while 22.5% mentioned the possibility of cheating in their venues. And finally, 11% opted for the no idea.

Table 6. Male and female PhD entrance exam candidates' satisfaction with proctors/inspectors

Item	Male			Female		
	Yes (%)	No (%)	Undecided (%)	Yes (%)	No (%)	Undecided (%)
Were the inspectors available at the exam session?	41.6	12.7	3.5	31.2	9.2	1.7
Did the exam proctors behave well?	50.9	3.5	3.5	34.1	1.2	6.8
Could the proctors keep the exam session peaceful?	45.7	7.5	4.6	33.5	5.8	2.9
Did the exam proctors provide the refreshments?	22	33.5	2.3	17.9	17.3	6.9
Was cheating possible at the exam?	13.9	38.9	9	8.1	28.3	13.7

41.6% of males in comparison with 31.2% of females stated that the exam inspectors were present at the exam sessions. While 12.7% of males and 9.2% of females mentioned that

the inspectors were not available. The Fisher's Exact Test Sig of .54 indicates that there was no statistically significant difference in the ideas of males and females in this regard.

Males were more satisfied with the behavior of the exam proctors comparing with females (50.9% vs. 34.1%). 6.8% females had no idea about the behavior of the proctors while this percent was 3.5% for males. The Fisher's Exact Test Sig value of .481 indicated no significant difference between genders.

Comparing the viewpoints of males and females, males (45.7%) were more satisfied with the performance of the exam proctors in keeping the exam session quiet than females (33.5%). However, there was no significant difference between genders (Sig = .53).

The higher dissatisfaction with the refreshments was for males (33.5%) in comparison with 17.3% for females. Totally, 22% of males, comparing with 17.9% of females, were satisfied with the refreshments and 6.9% of females had no idea in this regard. No significant difference was observed between the two genders in their satisfaction with the refreshments (Sig = .18). 38.9% of males in comparison with 28.3% of females stated that there was no cheating possibility in the exam session while 13.9% of males and 8.1% of females mentioned that cheating was possible. Moreover, 13.9% of females opt for the no idea. However, females and males did not differ in their idea about the possibility of cheating in their testing venues (Sig = .57).

The last research question was to find the general evaluation of the PhD exam candidates about the entire test administration process and if they have any recommendation for the exam administration improvement. 74.5% of the total participants of the study were satisfied with the entire process while 13.8% were dissatisfied. The following is a list of the most frequent comments proposed by the PhD exam candidates regarding the exam administration.

- The high possibility of cheating
- Better light, more comfortable seats
- Adding some extra time to the exam (at least 10 more minutes)
- Talking on the cell phone by some exam proctors
- More training for exam proctors
- Improving the commuting facilities
- Preparing enough parking lots
- Test takers' chatting at the exam session

5. Discussion

The purpose of this study was to see if the test administration conditions of the General English section of the PhD Entrance Exam were satisfactory from the exam candidates' perspectives. Standardized test administration, as one of the crucial steps in the testing cycle (Saville, 2012),

can be investigated from three different perspectives of timing, testing venue and the inspectors/proctors which were the focus of the first three research questions of this study. The last research question checked the general evaluation of the exam candidates about the entire testing process.

The issue of time was the focus of four items in the questionnaire which were asking for the suitability of the time, morning administration, evening administration and time delay in administration. The results indicated that 79% of the participants were satisfied with the administration time which was 8:00 a.m. 72% preferred to take the exam in the morning and finally, 60.1% percent of the participants stated that there was no time delay in the administration of the exam. The findings of this research question are in line with Monk (1990) who recommended administering the adult tests in the morning. The reason was that adults commonly perform better in the morning due to the temperature; therefore, the optimal performance time in a day is morning for adults. Moreover, referring to Wise et al. (2010), morning test administration yields more valid scores.

The reason for the superior morning performance might be explained from the cognitive memory functions and attention standpoints (Sjosten-Bell, 2005). Sousa (2001, p. 102) points out, that the “ability to focus in pre-adolescents and post-adolescents rises in the morning and then remains steady until about mid-day when there is a significant drop”. In other words, people’s minds are not busy, tired (due to daily activities) in the morning and therefore, they can function more drastically than any other times. According to Blake (1967), cited in Sjosten-Bell (2005), the basal arousal level continues to rise until 8:00 a.m. This might be the reason why the PhD exam candidates preferred to take the test at 8:00 a.m. Moreover, in examining attention, Muyskens and Ysseldyke (1998) investigated levels of attention of 122 students during a day. The findings showed that students were more attentively engaged during morning. Another factor might be the memory factor. Baddeley et al. (1970) found that short-term memory improves from early to mid-morning and then decreases steadily over the day. Likewise, the Folkard’s et al (1977) study proved that short-term memory recall was better in the morning than afternoon. Considering the multiple advantages of morning administration, it is recommended that the test administrators run the test, especially high-stakes ones, in the morning.

The second research question dealt with the testing venues issue. The factors that are of considerable importance are the location of testing venues, commuting, finding seats, ventilation and lighting of the venues (Tables 3 and 4). The results of the analysis indicated that 90.8% of the participants had enough information about the testing venues’ location; however, 75.1% of the participants could not easily commute to the testing venues of whom commuting to the testing venues was more difficult for males (44.5%). The majority of the participants, 90.8% could find their seats easily; the ventilation of the testing venues was satisfactory to 68.8% and finally, lighting was adequate from the perspective of 86.1% of the participants.

The findings of this research question are in line with what has been stated by Cohen and Wollack (2006) regarding the testing venue that if the testing venue is noisy, the test takers might be distracted and they cannot present their real competence in their test. Moreover, Shulman et al. (2011) found that venue and the measurement format affect the performance of test takers on a political knowledge test and finally, According to Douglas (2014, p. 54), “the room should have comfortable seating, ample desk or table space, comfortable temperature, good lighting and be reasonably quiet.... It is important to have an environment conducive to optimum test performance.” What can be concluded about the testing venues from the viewpoint of the PhD exam candidates is that they were satisfied with the testing venues in general; however, the main problem was commuting to the testing venues especially for male participants. The reason might be that the testing venues for the females are more accessible than those of males; therefore, commuting is easier for females than males.

The third research question sought to find the satisfaction of the PhD exam candidates with the exam inspectors/ proctors. Five questions (Tables 5 and 6) were checking for this issue, namely, the presence of the inspectors, satisfaction with the behavior of the proctors, peacefulness of the session, refreshments and finally the possibility of cheating. The general conclusions of the study regarding the inspectors/ proctors were that 72.8% of the participants stated that the inspectors were present at the testing venue to be referred to in the case of any need. 85% participants of the study were satisfied with the behavior of the exam proctors. However, males were more satisfied with the behavior of the exam proctors than females (50.9% vs. 34.1%). Considering the peacefulness of the exam session, 79.2% of the participants of the present study stated that the proctors could keep the exam session quiet enough.

Half of the participants of the study, 50.9%, were not satisfied with refreshments. The higher dissatisfaction with the refreshments was for males (33.5%) comparing with 17.3% for females. And finally, 66.5% of the participants mentioned that cheating was not possible in their testing venues due to the satisfactory performance of the exam proctors.

The matter of personnel is important to the extent that Fulcher and Davidson (2007) stated that the personnel or staff, as called by Saville (2012), needs to have training in how to deal with the test takers including checking their identity, giving instructions, managing the late arrivals, disruptive behaviors, those who finish early and especially cheaters. Besides, *standards for Educational and Psychological Testing* (AERA 1999) asserted that care must be taken to ensure the reliability of test scores through depriving the test takers of the opportunities to obtain scores by cheating or deceitful action. Also, since every deficiency in the administrative procedure may affect the validity and the quality of tests, it is recommended for the inspectors and visitors from the test development organization to inspect and monitor personnel and proctors (Saville, 2012) since, referring to Timmons et al.(2005), the administration or scoring of tests by personnel without the necessary qualifications is a serious ethics violation and may also

affect the validity of the results. In conclusion, Douglas (2014) recommended the exam proctors and personnel to behave politely and respectfully toward the test takers. All in all, the PhD exam candidates were satisfied with the issues related to inspectors/proctors; however, the only source of complaint was that they were not provided with the needed refreshments especially for males.

The final research question of the study was to check if the exam candidates were satisfied with the total test administration procedure and if they have any comments for improvement. Fortunately, 74.5% of the total participants of the study were satisfied with the entire process. Of the most frequent comments, three were related to the personnel, two were related to the timing issue and one was related to the testing venues.

6. Conclusion

The administration of language testing is of much more importance than it is usually dealt with in the language testing literature. In other words, it has been placed under the category of “test practicality,” but not treated in any detail, even if its significance to questions of validity has been duly acknowledged (Fulcher & Davidson, 2007). According to Fulcher (2010), getting the administration right is also about fairness to the test takers. Moreover, ensuring that these administration conditions are replicated for all test takers, so that no one experiences taking the test in a condition that provides either advantages or disadvantages, is a way of guaranteeing an unbiased test. As the results of the test administration investigation of the PhD Entrance exam indicated, in general, test administration conditions were satisfying for the exam candidates; however, there were some complaints about refreshments, commuting to the test venues, especially for male exam candidates. Therefore, refreshments and commuting issues should be cared and considered more.

References

- Alderson, J. C., Clapham, C. & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- American Educational Research Association (AERA), American Psychological Association (APA) and National Council on Measurement in Education (NCME) (1999). *Standards for Educational and Psychological Testing*. Washington, DC: AERA.
- Anderson, M. J., Petros, T. V., Beckwith, B. E., Mitchell, W.W., & Fritz, S. (1991). Individual differences in the effect of time of day on long-term memory access. *American Journal of Psychology*, 104(2), 241-255.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

- Bachman, L. F. & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bachman, L.F. & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
- Baddeley, A D., Hatter, J E., Scott, D., & Snashell, A. (1970). Memory and time of day. *Quarterly Journal of Experimental Psychology*, 22, 605-609.
- Brown, B., & McInerney, M. (2008). Changes in academic dishonesty among business students in the United States, 1999–2006. *International Journal of Management*, 25(4), 621–632.
- Cizek, G. J. (2003). *Detecting and preventing classroom cheating: Promoting integrity in assessment*. Thousand Oaks, CA: Corwin Press.
- Cohen, A. S., & Wollack, J A. (2006). Test administration, security, scoring and reporting. In R. L. Brennan (Ed.), *Educational measurement* (pp. 355-386). New York: American Council on Education/Praeger.
- Crown, D., & Spiller, M. (1998). Learning from the literature on collegiate cheating: a review of empirical research. *Journal of Business Ethics*, 17 (6), 683–700.
- Douglas, D. (2014). *Understanding language testing*. Routledge.
- Folkard, S. Monk, T.H., Bradbury, R., & Rosenthal, J. (1977). Time of day effects in school children's immediate and delayed recall of meaningful material. *British Journal of Psychology*, 68(1), 45-50.
- Fulcher, G. (2010). *Practical language testing*. London: Hodder Education.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. London and New York: Routledge.
- Gupta, S. (1991). Effects of time of day and personality on intelligence test scores. *Personality and Individual Differences*, 12 (11), 1227-1231.
- Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Kunnan, A. J. (2000). Fairness and justice for all. In A. J. Kunnan (Ed.) *Fairness and validation in language assessment: Studies in language testing* 9 (pp. 1–14). Cambridge: Cambridge University Press.
- Kunnan, A. J. (2003). Test fairness. In M. Milanovic & C. Weir (Eds.) *Select papers from the European Year of Languages Conference*, (pp.27–48). Cambridge: Cambridge University Press.
- Leigh, C., & Reynolds, C. (1982). Morning versus afternoon testing and children's intelligence test performance. *Perceptual and Motor Skills*, 55(1), 93-94.
- Mandinach, E B., Bridgeman, B., Cahalan-Laitusis, C., & Trapani, C. (2005). The impact of extended time on SAT® test performance. *ETS Research Report Series*, (2), i-35.
- McCabe, D., Treviño, L., & Butterfield, K. (2001). Cheating in academic institutions: a decade of research. *Ethics and Behavior*, 11(3), 219–232.

-
- Monk, T H. (1990). The relationship of chronobiology to sleep schedules and performance demands. *Work and Stress*, 4(3), 227-236.
- Muyskens, P., & Ysseldyke, J. E. (1998). Student academic responding time as a function of classroom ecology and time of day. *The Journal of Special Education*, 31(4), 411-424.
- Saville, N. (2012). Quality management in test production and administration. In G. Fulcher, & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 395-412). Abingdon: Routledge.
- Shulman, H. C., Boster, F. J., & Carpenter, C. J. (2011). *Do data collection procedures influence political knowledge test performance?* Paper presented at the annual meeting of the Midwestern Political Science Association in Chicago, IL.
- Sjosten-Bell, W. (2005). *Influence of time-of-day on student performance on mathematical algorithms*. <http://files.eric.ed.gov/fulltext/ED490712.pdf>. Accessed 3 May 2015.
- Sousa, D.A. (2001). *How the brain learns* (2nd ed.). Thousand Oaks, CA: Corwin Press, Inc.
- Timmons, J., Podmostko, M., Bremer, C., Lavin, D., & Wills, J. (2005). *Career planning begins with assessment: A guide for professionals serving youth with educational and career development challenges*. Washington, D.C
- Webb, W. B. (1982). Sleep in older persons: Sleep structures of 50- to 60-year-old men and women. *Journal of Gerontology*, 37, 581-586.
- Weir, C. (2005). *Language testing and validation*. London: Palgrave Macmillan.
- Wise, S. L., Kingsbury, G.G., Hauser, C., & Ma, L. (2010). *An investigation of the relationship between time of testing and test-taking effort*. Paper presented at the annual meeting of the National Council on Measurement in Education, Denver, CO.
- Xi, X. (2010). How do we go about investigating test fairness? *Language Testing*, 27(2), 147-170.
- Zieky, M., (2006). Fairness reviews in assessment. In S. Downing, & T. E. Haladyna, *Handbook of Test Development*. Mahwah, New Jersey: Lawrence Erlbaum Associates Inc., Publishers.