

The Sequence of Modules: A Facet in Language Proficiency Testing

Hesamoddin Shahriari Ahmadi

Department of English, Ferdowsi University of Mashhad

PO box 9177948974, Mashhad, Iran

Tel: 98-91-5510-1544 E-mail: hesam.shahriari@yahoo.com

Yesim Ozek

Northern Arizona University

English Language Department, Flagstaff 86001, AZ, USA

Tel: 1-928-380-0942 E-mail: yesim.ozek@nau.edu

Elif Demirel

Karadeniz Technical University

Department of English Language and Literature, 61080 Trabzon, Turkey

Tel: 90-462-377-3549 E-mail: elif@ktu.edu.tr

Received: December 21, 2011

Accepted: April 24, 2012

Published: June 1, 2012

doi:10.5539/ells.v2n2p11

URL: <http://dx.doi.org/10.5539/ells.v2n2p11>

Abstract

Given the widespread application of language proficiency examinations and their gate-keeping function, it is of utmost importance for researchers and test developers alike to identify and eliminate any facet of these tests which could be a potential source of invalidity or unreliability. One such facet is the sequence through which the four language skills are presented to the applicants. The present body of literature indicates that there has been no investigation into the order of skills in language proficiency tests. What is more, two established language proficiency tests, namely the International English Language Testing Service (IELTS) and the Test of English as a Foreign Language (TOEFL) present their skills in different sequences. This study sets out to determine whether altering the sequence of two skills on a language proficiency test (in this case, the IELTS) would result in any difference in applicants' performance on each individual skills. To this end, 120 learners of English as a Foreign Language (EFL) were asked to take part in two consecutive administrations of the IELTS, each time with a different sequence. The findings revealed that although intermediate and advanced learners performed equally well on both administrations, there was a significant difference in the performance of elementary learners across tests.

Keywords: Proficiency testing, Test method facets, IELTS, Sequence of skills

1. Introduction

Ever since the 20th century, there has been a constant growth in the number of people, crossing national and international borders for migration and educational purposes (Saville, 2006). The ability to effectively communicate with others in the classroom and work environment, not to mention social encounters, is mandatory for the survival and well-being of these immigrants, and also for their successful integration into the target cultures and communities. Considering what has been said, it is only natural for the countries and universities, receiving the immigrants, to establish criteria for language proficiency. However, due to the high-stake nature of the language proficiency tests utilized for gate-keeping purposes, it is upon the administrators of such tests to ensure the validity of their products and also the fairness of the decisions made based upon them. Failure to do so would seriously disadvantage many of the stakeholders involved in the testing process, including the administrators themselves. As a result, researchers, both internal and external, are encouraged to carry out studies in line with improving the validity and reliability of large-scale standardized tests. In fact, it is upon researchers and scholars to investigate the validity and reliability of language tests from a critical point of view and to identify and do away with any aspect of the test, which may be a source of unfairness or bias. Such sources may initially seem quite trivial on the surface, but upon closer, and not to

mention more critical scrutiny, one can understand the great impact they can have on the end results.

The International English Language Testing Service (IELTS) is a proficiency test, popular among non-native speakers of English, seeking education and work abroad. The IELTS is required by academic institutions and immigration offices in countries such as Australia, Canada, and the United Kingdom. Each year, applicants of this test spend huge amounts of money for preparation courses and materials, not to mention registering for the test itself. The least these candidates could expect in return for their time, effort, and expenditure, is to be fairly and validly assessed. Taylor (2000) states that upon closer analysis, we will see that in tests such as the IELTS, a complex community of participants and stakeholders are involved. The score received by an applicant will influence all of these stakeholders to a certain degree. Unfortunately, there are some aspects to proficiency tests such as the IELTS that have not been fully accounted for and lack a solid, research-based foundation. Each one of these potential sources of invalidity and unreliability could significantly impact the candidates' obtained score, and consequently affect their future life. These variations range from the overall format of the test (i.e. multiple choice, cloze, short answer, etc.) to seemingly trivial issues such as the font and size of the text used in the items.

According to Bachman (1990) the sequence, through which skills are tested, can have a considerable effect on the applicants' performance. The purpose of this study is to determine whether the sequence of testing skills on a language proficiency test could potentially influence the scores obtained by candidates on each individual skill, as well as the test as a whole. The design of this study consists of two distinct phases. The first phase aims to find out about the beliefs of applicants who are preparing themselves for partaking in the IELTS examination and whether they believe the sequence of skills plays a role in their performance. The second phase of the study investigates whether altering the sequence of skills on a proficiency test such as the IELTS would bring about any significant difference in the performance of candidates.

The determination of any given test method is based on a compromise between the desired authenticity, on the one hand, and logistical matters, on the other. The manner in which this compromise is achieved may adversely affect the fairness of the conclusions and decisions made based upon the results (McNamara, 2000). Mainstream language proficiency tests differ in the sequence in which they test the four main language skills. The main difference lies in the order of the listening and reading skills. The IELTS, for instance, begins with the listening module and then moves on to testing reading proficiency. Both the paper-based and Internet-based versions of the TOEFL, however, test reading prior to assessing listening. In Cambridge proficiency tests such as the FCE, CAE and CPE, reading precedes listening, with writing and use of English coming in between the two modules. This variation in sequence could either be due to disagreements over authenticity or ease of administration. In either case, the difference in sequence could possibly jeopardize the validity of inferences made on the basis of the results.

In order to investigate the validity of score-based interpretations made based upon a test, Bachman and Palmer (1996) asserts that we initially need to build a validation argument and then collect evidence in support of that argument. Validation arguments for a particular test will consist of claims and counterclaims about factors believed to influence performance on that test. In using a language test, the central claim is that the language ability we wish to measure is the chief cause of learners' performance. Many years of experience in the administration of language tests and countless studies carried out into this area bear witness to the fact that factors other than the intended language ability are also accountable for performance on language tests. Within a validation argument, other factors believed to affect performance should be articulated as counterclaims.

According to what has been said, in order to be able to examine the validity of score-based interpretations, we must distinguish between the effects of the abilities we wish to measure and the effects of other factors. In other words, according to Bachman's framework of validation, not only the ability of interest, but also other intervening factors predicted to effect test scores should be determined and defined.

Bachman (1990) provides a categorization of such factors and places them into one of the following broad categories:

1. *test method facets*
2. *test-taker characteristics, not part of the language abilities we wish to measure*
3. *random factors that are mainly unpredictable and temporary*

Among the three groups of factors, the test method facets are of particular importance in language testing. This is mainly due to the fact that the interaction between the characteristics of the test methods employed for eliciting test performance and the features of the language use context will directly influence the authenticity of the test and test tasks (Bachman, 1990). Hence, one could argue that a higher correspondence between the characteristics of the test method and the main features of language use contexts will result in a greater degree of authenticity of test tasks for

the applicants. The characteristics of test methods are controlled versions of contextual features of language performance.

The administrators of the IELTS are committed to enhancing the overall quality of their examination. Several systematic revisions have been made to the test over the course of its development and administration. In line with this openness to improvement, many studies have been carried out to investigate the effect of test method and its various facets on candidates' performance. Everett and Coleman (2003) explored the appropriateness of content and presentation of the listening and reading components. Studies have also been conducted into each of the four skills. For instance, the interpretation of prompts of the writing module was investigated by Mickan and Slater (2003). Brown (2000) and Shaw (2003) sought to determine whether the mode in which candidates were to present their writing (i.e., handwritten or typed) would influence scoring. For the speaking module, Develle (2009) examined the harshness/leniency and consistency estimates of examiner behavior.

A relatively small number of studies have been carried out into the issue of test organization. Kobayashi (2002) explored the effects of text organization and response format on the reading comprehension of 754 Japanese university students. The findings of that research revealed that both factors had a significant impact on students' performance. Kobayashi (2002) further argued that through well-structured tests, examiners can better differentiate between students with different levels of proficiency. This study and the conclusions drawn based upon it were later subjected to criticism by Chen (2004). The present study differs from that of Kobayashi's in that it looks into inter-modular organization of the test as opposed to the intra-modular organization or the sequence of sections and items within a skill or module.

2. Method

2.1 Participants

Each of the two phases of this study had its own group of participants. For the first part, a number of 60 participants, all of whom were familiar with the IELTS exam through attending preparation courses for the purpose of partaking in the examination at a private language institute, were asked to cooperate by responding to a questionnaire. From the total number of participants in this section of the study, 33 were female, and 27 were male. The mean age was approximately 30 with the youngest participant being 21 and the oldest being 48. As for the level education, three participants had a high school diploma, seventeen held a bachelor's degree, eight had obtained a master's degree and two held a doctorate degree. Among the respondents, seven had already taken part in the IELTS examination at a previous date.

The participants of the second phase consisted of 120 examinees. All participants had voluntarily agreed to take the examination and thus were not picked out subjectively by the researcher. Some of the examinees were undergraduate students of various majors, while others were EFL learners at various private language institutes. Due to the fact that almost all participants of this study were university students, studying at the undergraduate level, it could be assumed that all had, at the very least, undergone language instruction in high school. All participants were briefly introduced to the nature and purpose of the study and were told that they should take part in two examinations held on separate days. Using a pre-test, all examinees were divided into three groups of beginner, intermediate and advanced learners.

2.2 Instrumentation

2.2.1 Questionnaire

In order to find out about learners' opinions regarding various possible sequences of skills on the IELTS, a questionnaire was developed. The questionnaire initially asked the respondents about their manner of acquaintance with the exam and also some personal information such as age, and level of education. The items presumed that respondents were all familiar with the examination. In the second section of the questionnaire, all possible sequences of skills, including the standard order, currently put to practice, were listed for the respondents, who were asked to choose two orders, which they deemed to be most conducive to eliciting their best performance.

The options have been titled and presented below:

Sequence A: 1- Listening, 2- Reading, and 3- Writing

Sequence B: 1- Listening, 2- Writing, and 3- Reading

Sequence C: 1- Reading, 2- Listening, and 3- Writing

Sequence D: 1- Reading, 2- Writing, and 3- Listening

Sequence E: 1- Writing, 2- Listening, and 3- Reading

Sequence F: 1- Writing, 2- Reading, and 3- Listening

The options were randomly ordered on the questionnaire and no option was prioritized. Apart from selecting the two most favorable sequences, respondents were also asked to explain in a few sentences about the reason why they had chosen each specific order and why they believed it to be preferable to its alternatives. The questionnaire was validated through piloting prior to the actual administration.

2.2.2 EFL Tests

The study made use of two specimen language proficiency tests: the paper-based version of the Test of English as a Foreign Language (TOEFL), developed and administered by the Educational Testing Service (ETS), and the general module of the International English Language Testing System (IELTS), administered and managed jointly by the British Council, the University of Cambridge Local Examinations Syndicate (UCLES), and IDP Australia. Each of the tests was an officially released sample of the examinations, distributed by their respective administrators. From the IELTS, only the listening, reading, and writing modules were used. The present study did not concern itself with the speaking module of the examination, since this module is tested separately from the other three modules on a different date and therefore does not affect applicants' performance on the other skills. Two samples of the IELTS were administered through the course of this study and their equality in terms of difficulty was established through a pilot study.

2.3 Procedure

In the first phase of the data gathering process, the designed questionnaire was distributed among 60 learners of English as a foreign language, who were familiar with the format of the IELTS exam and its modules. Respondents were given ample time to read through the items and were encouraged to request clarification for any item they felt was ambiguous. From the provided responses, the two most frequently chosen sequences and the underlying reasons for why respondents had selected them as being more favorable were identified.

The two language proficiency tests were administered on separate occasions but at the same venue. The second test was held one week after the first. This interval was believed to be optimal, because learners' proficiency was unlikely to change significantly over this period of time. Also, had the interval been shorter than one week, participants might have still felt tired from the first exam. Great care was taken to provide similarly calm and un-distracting environments for all participants. Despite such attempts, all variables were not rigorously controlled by the researchers.

On the first administration of the IELTS, the skills were presented in the following sequence:

1. *Listening*
2. *Reading*
3. *Writing*

On the second administration, the sequence of the skills was altered, and they were presented as follows:

1. *Reading*
2. *Listening*
3. *Writing*

The scoring procedures used for the tests followed the guidelines set by the respective test developers: The TOEFL and IELTS papers were scored based on marking schemes developed and distributed by their respective developers. The listening and reading modules of the IELTS were objectively scored using the clearly defined answers provided in the marking schemes. As for the writing module, each essay was independently marked by two raters. Immediately before the scoring session, the raters underwent a rater training program (i.e., a norming session), consisting of an orientation to the writing test, a discussion of the IELTS writing band descriptors, and the sample scoring of a number of responses. Both raters were graduate students of TEFL and were experienced teachers of IELTS preparation courses.

3. Results and Discussion

Responses provided on the questionnaire were analyzed in order to find out whether sequence was in fact believed to be significant from the learners' point of view. The frequency table for the number of times each choice was selected has been brought below:

Insert Table 1 Here

As it can be seen, Sequence A, which was the currently established sequence by the administrators, was chosen

more than any of the other responses. Following this sequence, Sequences C (Reading, Listening, Writing), and B (Listening, Writing, Reading) were more frequently selected by the respondents, respectively.

In the questionnaire, the respondents were also asked to state their reasons and justifications for having selected a particular sequence. In this section, the reasons stated for each of the three most frequent sequences are presented. It is worth mentioning that because the responses were given in the respondents' L1 (Farsi), the most common explanations have been translated and reported below:

Sequence A:

1. The increasing need for attention in the writing module, which reportedly demands its placement at the end of the examination.
2. The relative ease with which listening items can be answered. This is said to give applicants a heightened self-esteem, helpful to completing the other modules.
3. The difficulty and need for attention often associated with the listening module, which could be better dealt with should it appear at the very beginning of the test.
4. Having the writing module appear at the beginning or the middle of the test reportedly causes anxiety since it is a time-consuming task and supposedly impairs the ability to respond to any other tasks which may follow.

Sequence C:

1. Having the reading module at the beginning of the examination has been said to increase the applicants' self-esteem, as this module is reported as being easier by some of the respondents.
2. Having not used up one's energy at the beginning of the examination, the respondents can allegedly complete the tasks in the reading module with more vigor and concentration.
3. Some respondents claimed that having the reading module at the outset of the examination helps the applicant become familiar with the test environment and reduce the existing levels of anxiety.

Sequence B:

1. Some respondents claimed that having listening as the first skill, will aid learners in regaining their self-esteem.
2. A few respondents believed that the more difficult skills ought to be left towards the end of the examination.
3. The listening module was said to require a great deal of attention. Therefore, it was claimed by some respondents, that having it at the beginning of the test does more justice to its nature. However, having the writing module, which was reported as being the easiest skill, allows the applicants to regain their energy for the somewhat more demanding reading module.

The responses obtained from the questionnaire made clear that sequences A, C, and B were respectively more popular among applicants of the IELTS who participated in this part of the study. This reveals that some applicants of the IELTS do in fact prefer to have the examination in a different order.

It needs to be pointed out that the popularity enjoyed by the standard sequence may be due to the fact that over time, applicants have become familiar with the current order of module presentation and have come to regard it as the best possible option. That is to say, this sequence was the only one the respondents were familiar with and had experienced first hand; and once presented with six possible options to choose from, they naturally opted for the standard sequence as at least one of their choices.

This cannot, however, be regarded as the only explanation for why the standard sequence was chosen more frequently than any of the other alternatives. The respondents themselves provided reasons as to why they had chosen each particular sequence. Their justifications for having chosen this choice chiefly dealt with matters of anxiety or self-esteem. They apparently chose a specific sequence because they either believed that it contributed to their self-esteem or influenced their levels of anxiety. Therefore, it could be claimed that applicants of this study, generally held the belief that the sequence for testing modules on the IELTS affected their performance through influencing their confidence and anxiety levels.

Respondents' explanations as to why they had favored a specific sequence over the other possible sequences followed a series of patterns. For instance, they showed general tendencies to place modules, which they believed to be more difficult, either at the beginning or at the end of the examination. For instance, some respondents claimed that since the writing module requires a great deal of their attention and is somewhat time-consuming in nature, it

had better be placed at the end of the examination. They believed that this would provide them with the opportunity to respond to the writing tasks with ease of mind, having already dealt with the tasks in the other two modules. Other respondents, however, preferred to have tasks, which they believed to be more difficult at the very beginning of the examination, arguing that they are far more attentive at that point and can better deal with such tasks.

Sequence C (Reading, Listening, Writing), selected as the second most popular sequence, was favored mainly due to having the reading module at the beginning. Almost all respondents who had selected this sequence referred to this point. According to the explanations provided, having the reading module at the beginning of the test boosted self-confidence and/or lowered affective filters. This, they claimed, was due to the relatively easy nature of the reading skill for them. In other words, participants stated that by having easier modules at the start of the test, they would become more prepared to deal with the demands of what they considered to be more difficult modules.

A second group of respondents who had selected Sequence C as one of their choices, also preferred to have the reading module at the beginning of the test, but for a completely different reason. This group of respondents claimed that the reading module demanded more effort on their behalf compared to any other one of the modules. Therefore, they argued that by having this module at the beginning of the examination, they could benefit from the high energy level available to them at the beginning of the test.

Once again, it can be observed that this sequence was also chosen due to the general tendency among applicants to prefer having relatively difficult or easy modules either at the beginning or at the end of the examination. This could possibly be accounted for by their belief that doing so would influence their self confidence and anxiety levels.

Finally, Sequence B (Listening, Writing, Reading), selected as the third most popular sequence, was preferred not only for having the listening module at the beginning of the test, as in the standard sequence, but also because having the writing module before the reading module allowed them some time to regain their lost energy and to move on to the reading section with lower levels of fatigue.

From what has been observed, it can be inferred that applicants differ in their sequential preferences depending on what skills they believe to be easier or more difficult for them, and also depending on whether they prefer to initiate the test with an easy or difficult module. It must not be forgotten, however, that applicants' beliefs do not always reflect what actually occurs on the examination itself, as we will see in the following phase of this study.

The second phase of this study investigated the effect of the sequence of listening and reading on applicants' performance on each skill as well as on their overall score. In this phase, two standard IELTS specimen tests were administered to 120 applicants who had already passed a 30-hour program designed to prepare them for partaking in the IELTS. All participants were screened for proficiency using the paper-based version of the Test of English as a Foreign Language (TOEFL). Using the pretest, applicants were divided into three groups of *elementary*, *intermediate* and *advance*. All participants were then asked to take part in two administrations of the IELTS with a two-week interval in between. The two-week interval was considered to decrease the chances of making significant gains in between tests or being affected by fatigue.

The results obtained by learners in the two administrations were compared using the matched t-test. The compared results included overall scores as well as scores obtained for the listening and reading modules. The purpose of such comparisons was to determine whether altering the sequence of reading and listening would result in a different band score either for each of the two skills or the test as a whole. The writing module was held constant as it was situated at the end of both administrations. Writing was considered while calculating overall scores but unlike listening and reading, which differed in sequence across the two tests, this module was not compared individually. In this study, the learners' level of proficiency was also of interest and therefore served as a moderator variable.

The table below summarizes the results of the comparison between the overall scores achieved by applicants in the two administrations. As can be seen, except for the elementary level applicants, there was no significant difference in the performance of the participants of this study across the two tests. In other words, altering the sequence of the reading and listening modules did not have a significant influence on the overall score of intermediate and advanced applicants; however, different sequences resulted in significantly different overall scores achieved by elementary participants.

Insert Table 2 Here

A second comparison was made between the listening scores obtained by each of the participants across the two administrations. The table below shows that intermediate and advance participants performed equally on both tests; however, elementary learners achieved significantly higher listening scores on the first test, in which listening preceded reading.

Insert Table 3 Here

Finally, the reading scores obtained in the first test were compared with those obtained in the second. According to the results, while there was no difference in reading scores obtained by intermediate and advance applicants, elementary participants performed significantly better on the reading module of the second test compared to the first. This is in contrast with what was previously observed with the listening module, where the same group of elementary participants attained significantly better results in the first administration. The results of the comparison between the reading modules can be seen in the table below:

Insert Table 4 Here

The sequence of skills in this study did not significantly influence the performance of intermediate and advanced applicants; elementary applicants, however, achieved better scores for the skill with which the test initiated. That is to say, in the first administration, they performed more successfully on the listening task; while in the second examination, they achieved better scores on the reading module. Based on the findings of this study, it could be claimed that the sequence of skills systematically affected the results of elementary learners.

Further studies would have to be carried out to identify the various possible reasons underlying this systematic influence. However, opinions elicited from applicants in the first phase of this study could also shed some light on this matter and possibly provide some clues for the identification of causes. As previously discussed, learners claimed to have higher levels of energy and attention at the beginning of the test. This was the reason why most of them chose to start the examination with skills they perceived as being more difficult or demanding more attention. This account could possibly explain the data found in the second phase of the study. If at the beginning of the examination, learners, as they themselves claim, do in fact benefit from greater vitality and available attention span, the more desirable performance of elementary applicants on the initiating module could be explained in light of these two intuitively determining variables. The question which remains is why intermediate and advanced applicants, given the fact that they were facing the same conditions and difficulties, did not perform differently on the two administrations. If stamina, vitality and attention resources are greater towards the beginning of the examination, one would expect these factors to influence the performance of all learners and not just those with lower levels of proficiency.

Attention is defined as the process through which one encodes language input and keeps it active in one's short-term memory and/or retrieves it from one's long-term memory (Robinson, 2003). Controlled L2 processing, which most elementary level learners draw upon, is said to be more attention-demanding compared to automatic L2 processing which is more common among intermediate and advanced learners (DeKeyser, 1997, 2001). According to such a view, elementary level applicants are more in need of attention resources than their more proficient counterparts who deal with language through automatic processing. This could possibly explain why elementary learners performed better on the first skill of the examination (regardless of whether it was the listening or reading skill), while intermediate and advanced learners performed equally on both examinations since they managed to save up on enough attention resources through automatic language processing abilities. It is important to mention, however, that studies would need to be designed and implemented in order to determine the degree to which attention affects the performance of learners on various skills and at various points in time during an examination.

4. Conclusion

The evaluation of the data derived from this study provided key information regarding the candidates' beliefs about the sequence of testing skills on proficiency tests such as the IELTS. The study also provided information about the effect of the sequence of module presentation on applicants' performance on each of the modules as well as their overall score. The findings reveal that test takers generally believe that the sequence of modules could affect their performance through influencing their self-confidence, attention and anxiety. Applicants of the IELTS were also shown to prefer having modules which they believed to be easier or more difficult either at the beginning or at the end of the examination.

The findings also revealed that elementary learners performed better on tasks which appeared at the beginning of the test. It was also speculated that a possible cause for this improved performance could be the greater attention resources available to applicants at the beginning of a test. This implies that proficiency test designers and administrators should be aware of the influence the sequence of skills can have on their candidates' performance on each individual skill and even possibly on the overall score.

One might argue that if according to the findings of this study, any skill placed at the beginning of a test were to elicit better performance by applicants, the decision as to which skill should take this position would be largely arbitrary. This is because by placing one skill at the beginning of the test, performance on another skill would be compromised. Future studies would need to be conducted to determine the extent to which these results would hold true in the case of other proficiency tests and other language-testing contexts. In addition, the possible effect of a

break in between skills could also be examined. If attention and vitality do in fact give rise to the difference in performance, as claimed by the applicants themselves, having an interval in between skills could help elementary learners regain their resources and perform their best on the examination. In conclusion, administrators and test designers ought to consider the sequence of skills as an important facet when designing and validating their language tests and attempts should be made to use the sequence, which not only causes applicants to perform their best on each individual module or skill, but also functions as a fair instrument in measuring candidates' language proficiency.

References

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: OUP.
- Bachman, L. F., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Brown, D. H. (2000). *Principles of language learning and teaching* (4th Ed.). London: Pearson Education.
- Chen, L. (2004). On text structure, language proficiency and reading comprehension test format interactions: a reply to Kobayashi, 2002. *Language Testing*, 21(2), 228-234. <http://dx.doi.org/10.1191/0265532204lt281xx>
- DeKeyser, R. (1997). Beyond explicit learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition*, 19, 195-221.
- DeKeyser, R. (2001). Automaticity and automatization. In P. Robinson (ed.), *Cognition and second language instruction* (pp. 125-151). Cambridge: Cambridge University Press.
- Develle, S. (2009). Certificating IELTS writing and speaking examiners. *Research Notes*, 38, 26-29.
- Everett, R., & Colman, J. (2003). A critical analysis of selected IELTS preparation materials. *IELTS Australia Research*, 5, 20-29.
- Kobayashi, M. (2002). Method effects on reading comprehension test performance: text organization and response format. *Language Testing*, 19(2), 193-220. <http://dx.doi.org/10.1191/0265532202lt227oa>
- McNamara, T. (2000). *Language Testing*. Oxford: Oxford University Press.
- Mickan, P., & Slater, S. (2003). Text analysis and the assessment of academic writing. *IELTS Australia Research*, 4, 40-52.
- Robinson, P. (2003). Attention and memory during SLA. In Doughty, C. J., & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 631-678). Oxford: Blackwell Publishing. <http://dx.doi.org/10.1002/9780470756492.ch19>
- Saville, N. (2006). Language testing for migration and citizenship. *Research Notes*, 25, 2-4.
- Shaw, S. (2003). Legibility and the rating of second language writing: the effect on examiners when assessing handwritten and word-processed scripts. *Research Notes*, 11, 7-10.
- Taylor, L. (2000). Stakeholders in language testing. *Research Notes*, 2, 2-4.

Table 1. Frequency table for questionnaire responses

	Sequence A	Sequence B	Sequence C	Sequence D	Sequence E	Sequence F
f	31	20	23	15	15	16
percent	35%	16.6%	21.6%	8.3%	8.3%	10%

Table 2. Paired-sample T-test results comparing the overall performance of participants from three groups (Beginner, Intermediate and Advanced) across the two administrations

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	OverE1 OVERE2	2.77	3.59	.56	1.62	3.92	4.88	39	.00
Pair 2	OverI1 OverI2	-.41	1.87	.29	-1.01	.18	-1.39	39	.17
Pair 3	OverI1A OverI2A	.33	1.88	.29	-.26	.93	1.13	39	.26

Table 3. Paired-sample T-test results comparing the performance of participants from three groups (Beginner, Intermediate and Advanced) on the listening module across the two administrations

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	L1E - L2E	4.85	2.60	.41	4.01	5.68	11.76	39	.00
Pair 2	L1I - L2I	-.27	1.37	.21	-.71	.16	-1.26	39	.21
Pair 3	L1A - L2A	.27	1.10	.17	-.07	.62	1.56	39	.12

Table 4. Paired-sample T-test results comparing the performance of participants from three groups (Beginner, Intermediate and Advanced) on the reading module across the two administrations

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	R1E - R2E	-2.10	1.86	.29	-2.69	-1.50	-7.12	39	.00
Pair 2	R1I - R2I	-.12	1.22	.19	-.51	.26	-.64	39	.52