

## Arabic Language Testing in Iran: A Content Analysis

Mehdi Zamani<sup>1\*</sup>

---

1. Department of Islamic Law and Jurisprudence, Izeh Branch, Islamic Azad University, Khouzestan, Iran.

\* Corresponding Author's Email: mehdizamani58@yahoo.com

---

**Abstract** – It is generally a given in critical language testing that countries with central educational systems enact their preferred language policies through language tests (Shohamy, 2001). To shed some light on how language policies regarding Arabic have changed, or rather have remained stable, this article traces the quantitative changes that have taken place in Arabic language testing in Iran over the last three decades. Given the centrality of the decisions made based on the scores on high stakes Arabic tests in the lives of Iranian students, it is of prime importance for the tests to be designed and items be written to a well justified test specification (Fulcher, 2010). To examine whether this basic requirement of test design is observed I content analyzed one such test to examine the extent to which it is aligned with the established priorities of the field. The corpus of tests under investigation in this study consisted of three sample tests, each of which belonging to a decade in the time span of the last three decades. Results suggest that the importance attached to Arabic language has been on the increase, though not in a statistically significant way.

**Keywords:** Arabic, Language, Tests, High stakes

---

### I. INTRODUCTION

Scores on summative and externally administered tests of Arabic are used to make a number of crucial decisions about Iranian citizens. Of the former, Arabic test scores constitute part of the aggregate score which is used to determine whether students are qualified to move to the next educational level and which ones should be denied the opportunity. As such, Arabic scores are at play in making progress decisions twice in secondary education; once by the end of junior high school and a second time at the end of senior high school. The latter gate keeping function is even more crucial because Arabic scores would be used to determine graduation from high school. The story does not end here, however. Arabic features in all Universities Entrance Exams (UEE) regardless of school graduates' field of study. Together with Persian literature, English and theology, Arabic forms the component of the UEE that is common for all test takers no matter what their past or future fields of study are. Arabic test items carry the same weight as that of core subjects like chemistry, math, and biology for many prospective college students. Considering this ever presence of Arabic tests at different stages of the educational careers of Iranian students, it is difficult to overestimate their importance as a worthwhile subject of inquiry.

Given the centrality of the decisions that are made based on the scores on high stakes Arabic tests in the lives of Iranian students, it is of prime importance to make every effort to ensure that such tests are designed in accordance with the latest knowledge and technology of the field. To my knowledge, however, no academic study has addressed this issue as of yet. To help trigger a debate on this issue, this study seeks to trace the developments that have occurred in the content and construction of Arabic language tests over the course of the last three decades. In particular, I will focus on constructs tested and test method facets of the tests and the changes they have undergone through over the specified time.

## II. REVIEW OF LITERATURE AND METHOD OF STUDY

Language tests are as old as the hills. The recorded history of mankind abounds with stories in which language tests in one form or another have been used for a diverse range of purposes. In the absence of modern proofs to identity like passports and social security numbers, our early ancestors were left with no further option other than language to determine the identity of friends from those of foes. History has it that a language test as brief as the pronunciation of one word served as the basis for life and death decisions in part of Europe that is now called the United Kingdom. The notorious word was *shibboleth* and depending on whether those involved could pronounce the first letter as /s/ or /sh/ their fate would be sealed (McNamara, 2000; McNamara & Rover, 2006). Similar incidents have been cited in other parts of the world including the middle east, far east, the southern hemisphere and elsewhere (McNamara & Rover, 2006).

The modern day language tests, however, though remain similar to ancient ones in function, have changed in scope and extent of sophistication, largely thanks to the standardization movement in testing, which is believed to have been pioneered in China during the era of Hann dynasty (Weir, 2005). However, the first modern language tests did not appear until nineteenth century, when Cambridge University created a test of English language which comprising of an essay, a literature section and some translation from and into French and German (ibid). On the other side of the Atlantic, language testing was developing by ETS with a preoccupation with consistency and reliability, in stark contrast with the practice in Cambridge, which was mainly centered on validity.

Since the pioneering efforts of Cambridge and ETS, language testing has evolved into a huge multibillion industry, with millions of test takers around the globe sitting IELTS and TOEFL annually. In tandem with developments in educational measurement, language tests have improved in significant ways since Cambridge pioneering English tests. Though ETS and Cambridge ESOL remain the leading forces in this market, language tests are today produced and administered by many other national and international institutions around the world. Moreover, although the bulk of literature in language testing is on measuring English proficiency, the testing of other languages has benefited from innovations and insights in English language testing, Arabic supposedly being no exception.

As I left my comfort zone, English language testing, to venture into mysterious world of Arabic language testing I found out that this is literally an uncharted territory. More

pronounced is the paucity of empirical research on Arabic testing in Iran, a Google Scholar search of which yields almost next to nothing. Interestingly, the one pertinent study carried out on the testing of Arabic has been published in *Language Testing*, the focus of which is almost exclusively on English testing. In their study, Shohamy, Donitsa-Schmit, and Ferman (1996) compared the washback effects of Arabic tests with those of English tests administered in Israel's educational system. To collect data, they administered questionnaires, carried out interviews, and analyzed documents. They found that slight modifications in the content and methods of English tests brought about significant changes in the way teachers taught English, in the way learners learned, and in how courseware were compiled and authored. Parallel washback effects were not found for Arabic tests. Shohamy et al. (1996) argue that such stark differences in the impacts the two sets of language tests trigger is to be accounted for by considering the status of the two languages and the uses that are made of the results of the Arabic and English tests. They further reason that in Israel English is seen as a language of modernity the learning of which improves learners' educational and career choices, whereas Arabic does not enjoy such a prestige.

Aside from what was reviewed above, the literature on the numerous Arabic tests given annually to Iranian students is almost nonexistent. Safari (2014) is perhaps the only exception, though it cannot be qualified a serious academic study since it aimed at helping test takers know which sources to rely more on rather than to shed light on the mechanism of Arabic testing. To make a modest contribution to this slim literature, I set out to seek answers to the following questions:

1. How has the status of Arabic language altered over the three specified decades as reflected to the number of test items allocated to high stakes Arabic tests?
2. What is the implicit test specifications to which Arabic language tests are written?

Given the nature of this inquiry, content analysis approach was adopted. However, considering the enormity of the corpus of all the Arabic tests that have been produced over the last three decades by Sazeman Sanjesh, inspecting the entire body of the corpus was not feasible; not it is basically conventional in research, making sampling inevitable. Thus I sampled from all the tests given to this day, leading us to a limited corpus comprising of three tests, each decade being represented in the sample by one test. More specifically, the corpus body consists of one Arabic test given in the seventies, one in the eighties, and one in the nineties (years are according to the Persian calendar).

### III. RESULTS

Table 1 gives the quantitative changes made to Arabic tests over the last three decades. Obviously, the number of test items has been on the increase through the years, though in an incremental fashion. While in the seventies there were twenty Arabic test items, this number reached 25 in the last decade. The increasing number of test items could have had either a language policy or a technical justification (more on this in the next section).

Table 1. *Number of test items in each Arabic test over the three decades.*

decade	Number of Test Items
Seventies	20
Eighties	22
Nineties	25

The above judgment of the figures in the table is based on eyeballing the table, which is not necessarily a principled way of approaching statistics. To decide whether the differences in the number of test items in each year, is not a mere coincidence or due to chance fluctuations, I need to run rigorous statistical tests. As such, table 2 illustrates the results of Chi Square test run on the data in table 1.

Table 2. *Chi Square test results*

Test Statistics	
Number of test items	
Chi-square	.567a
df	2
Asymp. Sig.	.753

As the results from table 2 indicates, the eyeballing impression of the figures in the table is not short of misleading because the critical p value turns out to be .75, meaning that there is 75 in every one hundred likelihood of this result being a random event, taken place out of a random coincidence. Caution, however, has to be exercised even with this statistical analysis because of the limitations in my sampling. Analyses based on a larger sample may well attest to the impression I get from eyeballing the table. What can be said of this data with a large margin of error is that sweeping conclusions are difficult to draw either in the direction of the eyeballing impression or in favor of the statistical analysis.

Uncovering the test specifications to which Arabic tests are written was the second concern of this study. To that aim, I analyzed the content of a recent retired version of Arabic exam. Table 3 gives an account of the outcome of the content analysis.

Table 3. *The number of Items for each category in the Arabic exam*

Category	Number of Items
Syntax	10
Conjugation	3
Reading Comprehension	4

Translation from Persian to Arabic	2
Translation from Arabic to Persian	6
Total	25

As the table shows, syntax receives the highest number of items (10). Three items gauge test takers' knowledge of conjugations, itself being a subcategory of syntax. Taken together, it is inferred that syntax dominates the test, with almost half of the test items given to it. The next category in the order of number of items is *translation from Arabic to Persian*, with six items. A couple of items tap into test takers' mastery of translation from the source language, Persian, to the target language of Arabic. Translation from and into Arabic constitutes one third of the total number of test items, indicating that in Arabic testing translation competence is considered an essential construct in test takers' repertoire of constructs. Mention should be made of the fact that the kind of translation expected of students is one that is strictly faithful to the original. The following is an example.

من عرف الدنيا معرفه حقيقيه، لا تخدع قلبه ظواهرها الخلابه  
الف. اگر کسی واقعا دنیا را بشناسد، ظاهر دلربایش او را فریب نمی دهد  
ب. آنکه حقیقت دنیا را شناخته باشد، ظاهر فریبنده آن فریبش نمی دهد  
ج. آن کسی که عمیقا دنیا را شناخت، به ظواهر خوش آن دل خوش نمی کند  
د. کسی که دنیا را به طور واقعی بشناسد، ظواهر دلربای آن قلبش را نمی فریبند.

As the above item demonstrates, all the four alternatives almost gives the same idea but in different wordings. Thus, if one is concerned with translating the gist, free translation, all alternatives can be taken as true. But since test designers appear to be in favor of a strict word for word translation, the option closest in form to the root sentence, the last one, is considered the correct answer and others are considered distractors. To get the item right, students need to draw on the syntactic nuances which set the alternatives apart. Put it another way, grammar is intimately implicated in these translation items too, showing that the model of language ability the test designer ascribe to is predominantly comprised of grammar and rules, which reminds us of structural linguistics.

Reading comprehension is also measured using a few items, compared to the categories of syntax and translation. Overall, there are four items targeting the construct of reading comprehension. It should, however, be pointed out that reading comprehension passages are rather short ones, not long stretches of discourse.

#### IV. CONCLUSION

A mere glance at the number of questions in each test over the decades indicates it has followed an ascending trend. This might have be for any or both of the following

possibilities. The first possibility, which has less to do with the testing per se and more to do with language policy, is that policy makers have attached more importance to Arabic and the more space and resources allocated to Arabic tests is a reflection of that increased importance. The other possibility is the technical one, having to do with test reliability. It is generally believed that the more items in a test the more reliable it would be (see Bachman, 2004). However, this possibility cannot be verified in the absence of solid statistical evidence comparing the reliability index of each of the tests. Such evidence is by now missing to my knowledge. Thus, the increase in the number of Arabic items in universities entrance exams could be a language policy issue, could be about improving the technical qualities of the tests, or might be for both concerns. Rejecting or verifying any of the aforesaid possibilities invites separate investigations.

Therefore, it is inferred that the test specifications to which Arabic tests are written is heavily informed by the Structuralist School of linguistics, as is reflected in the large number of items that demand test takers to master rules of syntax and conjugation. Obviously, Arabic test designers define language competence as being primarily a matter of learning the rules of language. Current thinking in language testing does not approve of such definitions of language. Rather, it is regarded outdated and out of sync with modern theories of language ability and linguistics. However, caution should be exercised as this conclusion is an inference based on test items. What goes on in the mind of test designers and policy makers may not be a linear antecedent to what is embodied in the test. Rather, this invites for further enquiry to probe Arabic item writers for their understanding of the construct of language ability.

## REFERENCES

- Bachman, F. L. (2004). *Statistical analyses for language assessment*. Oxford: Oxford University Press.
- Fulcher, G. (2010). *Practical language testing*. Hodder Education
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press
- McNamara, T., & Roever, C. (2006). *The social dimensions of language testing*. Hodder Education
- Safari, K. (2004). Sources of Arabic tests. *Nourmags*
- Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing*, 18, 27-65
- Shohamy, E., Donitsa-Schemit, S., & Ferman, I. (1996). Test impact revisited: washback effect over time. *Language Testing*, 13(3), 298-317
- Weir, C. (2005). *Validity in language testing: An evidence-based approach*. Oxford: Oxford University Press