



Applying EALTA Guidelines as Baseline for the Foreign Language Proficiency Test in Turkey: The Case of YDS*

Nurdan Kavaklı^{a *}, Sezen Arslan^b

^a Hacettepe University, Ankara, Turkey

^b Yüzüncü Yıl University, Van, Turkey

Abstract

Within the scope of educational testing and assessment, setting standards and creating guidelines as a code of practice provide more prolific and sustainable outcomes. In this sense, internationally accepted and regionally accredited principles are suggested for standardization in language testing and assessment practices. Herein, ILTA guidelines for good practice proposed by International Language Testing Association (2007), ALTE code of practice by Association of Language Testers in Europe (1994), JLTA code of good testing practices by Japanese Language Testing Association (2002) and EALTA guidelines for good practice by European Association for Language Testing and Assessment (2006) can be cited. Amidst them, the EALTA guidelines have been adopted to ‘frame a validity study’ (Alderson, 2010: 63) for language testing and assessment practices. In this sense, due to the abundance of guidelines and principles, it is expected to see myriad of practices to be well-implemented and documented. However, documentation on aforementioned practical cases is rare with a few empirical studies conducted (Alderson & Banerjee, 2008; Alderson; 2010; De Jong & Zheng, 2011). Accordingly, in this paper, a practical case study on YDS (foreign language exam in Turkey) is applied regarding the EALTA guidelines with a special concern on the development of tests in national and/or institutional testing units or centers. It is, therefore, aimed to tackle the question whether YDS adheres the principles purported by EALTA with its probable high-stake consequences. Thus, the results have indicated that taking the EALTA guidelines in the course of the test development process as baseline promotes value-added language testing and assessment practices.

© 2016 IJCI & the Authors. Published by *International Journal of Curriculum and Instruction (IJCI)*. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (CC BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: EALTA Guidelines; YDS; language testing; educational assessment

1. Introduction

As a part of educational testing, language testing serves to provide standards and/or guidelines as baseline values. Regarding the mission stated within the Guidelines for

* Corresponding author : Nurdan Kavaklı
E-mail address: nurdankavakli@gmail.com

good practice in language testing and assessment, EALTA aims to nourish the perception of the theory behind language testing and assessment practices throughout Europe. By publishing the Guidelines (EALTA, 2006) and making it available in more than thirty languages, it is aimed to bolster general principles which are to be verified during the test development process. In the same vein, EALTA presents a rationale for the improvement of sharing of aforementioned practices throughout Europe by those guidelines set. Creating a sphere by the inclusion of diversity in education systems and assessment traditions, EALTA addresses three different types of audiences advising accountability, transparency and quality improvement for testing and assessment practices. These audiences can be listed as those who are involved in (1) the training of teachers in testing and assessment; (2) in-class testing and assessment; and (3) test development in national or institutional testing units or centers.

However, when the practicality is concerned, there is the scarcity with a few empirical studies conducted (Alderson & Banerjee, 2008; Alderson, 2010; De Jong & Zheng, 2011) on the application and documentation of the EALTA Guidelines for goodness in language testing and assessment practices. To elaborate, Alderson (2010) and Alderson & Banerjee (2008) devised a survey questionnaire to the English test providers who are labelled as the third type of audience above mentioned. De Jong & Zheng (2011) checked Pearson Test of English Academic in relation to the third type of audience, as well. Herein, seven basic concepts targeted for the third type of audience have been touched upon for these studies conducted. These concepts can be labelled as: (a) test purpose and specification; (b) test design and item writing; (c) quality control and test analyses; (d) test administration; (e) review; (f) washback; and (g) linkage to the Common European Framework of Reference for Languages (hereafter CEFR). As a result, the authors argued that the EALTA guidelines could be adopted to ‘frame a validity study’ (Alderson, 2010: 63) in relation to the use of guidelines for language testing and assessment practices. From that point of view, considering the third type of the audiences listed above, it is aimed within the scope of this study whether YDS, a language proficiency test administered in Turkey, adheres the aforesaid seven concepts purported by EALTA with its probable high-stake consequences. Therefore, a thematic analysis is conducted in order to answer whether the language proficiency test administered in Turkey complies with the aforementioned seven concepts noted by EALTA.

1.1. The Language Proficiency Test Administered in Turkey

The YDS, which is formerly known as KPDS, is a language proficiency test administered by OSYM (Measuring, Selection and Placement Center) as the body which is responsible for organizing large-scale examinations at national level in Turkey (Öğrenci Seçme ve Yerleştirme Merkezi, 2013a). The test, which can now also be taken electronically, is conducted every six months, one in the fall and the other in the spring

terms. The test is administered in several languages, albeit mainly in English. To note, the test in spring term is designed in more than twenty languages such as German, Chinese, French, Japanese, Persian, Spanish and the like whereas the test in fall term is designed solely in Arabic, French, English, Russian and German.

To probe into its nature, the test consists of 80 multiple-choice question items for which the test takers are given 180 minutes to finish. These items mainly deal with English vocabulary, grammatical structures, English-Turkish translation, Turkish-English translation, sentence completion, paragraph completion, dialogue completion, restatements, reading comprehension and odd-one-out on the irrelevant sentences in a paragraph. However, in some other languages such as Danish, Armenian, Greek, Chinese, Japanese and Korean, the test is composed of merely translation items from Turkish to the target language and vice versa. For the multiple-choice question items, each item is 1.25 points to score, whereas the translation items on merely some other languages noted above are evaluated by an academic jury in OSYM.

As well as defining one's own foreign-language skills, the test is taken by any governmental employee who is working as a civil servant, military personnel or an academic to be paid extra money based upon the scores they get. The test is also used to appoint employees to the positions abroad. Moreover, the test is taken for the enrollment in universities for master and PhD programs in Turkey, for the entry to some undergraduate courses and for the workplace opportunities provided. One more to add, the test is taken by the high-school students, who desire to become an English teacher, for the enrollment in bachelor's degree at universities.

In the light of these, the high-stake nature of the YDS pins down the exploitation of international standards. However, as the YDS is administered and approved within Turkey merely, it could not be said that it is accredited internationally around the world. Accordingly, the YDS has been analyzed thematically within years. In his qualitative study, Özmen (2011) examined the washback effects of UDS (which is also a similar language proficiency examination administered by OSYM) on prospective academics in Turkey. As a result, he reported that the test had some negative washback effects on some micro- and macro-level variables such as cognitive learning, course and material expenses, and L2 competences. Similarly, Yavuzer and Göver (2012) investigated the opinions of the academics on foreign language examinations such as KPDS and UDS and their language proficiency levels. Their study mushroomed that the academics seemed to perceive these tests as a set to further academic development as four basic language skills were not equally considered within. In another study, Güleç (2013) probed into how academics were studying YDS together with their attributions of success and failure, and their overall opinions towards the test. Accordingly, it was reported that though academics had a positive belief towards learning English, they took YDS just because they were obliged to do so for their ongoing career as an academic.

In another context, Kıray (2015) made a macro-structure analysis of reading comprehension paragraphs of the exam within the circulation of 2003-2013. Demir and Genç (2016) reported an analysis of the KPDS, as the former of version of YDS, based on the Common European Framework of Reference for Languages. Accordingly, the equivalency tables declared by the Council of Higher Education (YOK) and OSYM were interpreted regarding the levels projected by the Framework. On the other hand, Akın (2016) investigated YDS within the scope of adult education and language for specific purposes. It was reported by the findings of this study that though it seemed as an advantage for test takers to be familiar with the question types thanks to the originality of the questions and the similarity of the test format, the main downside of the test was not assessing all of the four basic skills within. Besides, Külekçi (2016) made a concise analysis of the foreign language examination, YDS, in Turkey with its possible washback effects. Herein, the construct of the test was analyzed in respect of assessing language proficiency thoroughly. However, it is found out that there is a gap in the literature that provides a review of the application of any European language testing and assessment criteria for this exam.

Therefore, in this paper, considering the third type of the audiences listed above, which is test development in national or institutional testing units or centers, it is aimed within the scope of this study whether YDS, a language proficiency test administered in Turkey, adheres the seven tenets purported by EALTA with its probable high-stake consequences. In this context, a thematic analysis is conducted in order to answer whether the language proficiency test administered in Turkey comply with the aforementioned seven concepts noted by EALTA.

2. Purpose and methodology of this study

YDS is a language proficiency test which is administered by OSYM. As it is a widely used test in order to measure the language proficiency in Turkey, it may be of enormous importance to analyze the test through guidelines which foster standardization in language testing. Therefore, this study aims to investigate YDS in terms of EALTA principles which are employed for test validation. Thematic analysis is applied in order to find out whether YDS adheres the principles of EALTA. Sample YDS tests, the information on OSYM website, test application guides, test specifications offered by OSYM and equivalency table of YDS are utilized in order to perform thematic analysis of the test.

3. Results

3.1 YDS in the context of EALTA Guidelines

The following sections are aligned by the seven aforementioned tenets. Answers to each question are inscribed under the subheadings, which are laced with specific examples and documents to support the conditions in which the guidelines are applied.

3.1.1 Test purpose and specification

This section embraces the test purpose of YDS and implementation of test specification in test development process.

3.1.1.1. How clearly is/are test purpose(s) specified?

The purpose of YDS is to measure general foreign-language proficiency of the test-takers. The test purpose is explicitly and clearly stated by the administrative body of the exam (Öğrenci Seçme ve Yerleştirme Merkezi, n.d.-a).

3.1.1.2 How is potential test misuse addressed?

To avoid misinterpretation of the results and any potential misuse, OSYM asserts a table comprised of five levels out of 100 as the total score. Accordingly, the scores between 90-100 are the indicators of A level whereas those between 80-89 sign B level. To add more, the scores between 70-79 indicate C level while those between 60-69 point to D. Finally, the scores between 50-59 refer to E level. This classification of scores is available online to aid test takers as it sets standards of scores to be used for required purposes (Öğrenci Seçme ve Yerleştirme Merkezi, 2016a).

3.1.1.3 Are all stakeholders specifically identified?

Test stakeholders are labelled as test takers and test score users, which are directly signaled by the test specification document.

3.1.1.4 Are there test specifications?

Following the decisions set on the test purposes, construct and content that are planned to be measured, the relevant people, who are actually the members of the test development team, are expected to come up with a test specification document. Herein, test specifications portray the purposes, constructs, domains, content, length, context, test takers' characteristics, conditions and procedures required for the implementation phase, scoring criteria and reporting test scores.

3.1.1.5 Are the specifications for the various audiences differentiated?

Test specifications which guide the development and employment of YDS do not seem to be differentiated for the various audiences.

3.1.1.6 Is there a description of the test taker?

The desired population of test takers is described as those who are required to provide evidence of their foreign language proficiency skills either to study at universities or to use it for professional career at work (Öğrenci Seçme ve Yerleştirme Merkezi, n.d.-a)

3.1.1.7 Are the constructs intended to underlie the test/subtest(s) specified?

The construct that YDS propounds is to assess mainly receptive skills as there is no item dealing with language production such as speaking and writing. In the light of these, the questions primarily deal with vocabulary items, grammatical forms, sentence structures, translation, reading comprehension, dialogue completion, restatement and odd-one-out structures (Öğrenci Seçme ve Yerleştirme Merkezi, 2017)

3.1.1.8 Are test methods/tasks described and exemplified?

The test items are composed of 80 multiple-choice questions. There is neither open-ended nor short-answer type of questions within. Some sample questions from above-mentioned areas of interest are given online by OSYM (Öğrenci Seçme ve Yerleştirme Merkezi, n.d.-b).

3.1.1.9 Is the range of student performances described and exemplified?

The test takers' performances are described by means of a scale composed of 5 levels from A to E. This classification appears to be local because test-takers' scores can mostly be interpreted within the context of Turkey; however, YDS scores which are accepted as the equivalences of such tests such as TOEFL and IELTS could be found on OSYM's website. (Öğrenci Seçme ve Yerleştirme Merkezi, 2013b).

3.1.1.10 Are marking schemes/rating criteria described?

The marking scheme for YDS scores is described in detail. The rating for each question is 1.25 points. Only correct answers are considered and the wrong answers do not have any additional decrease unlike the exams conducted by OSYM, in which every 4-wrong-answer takes 1 correct answer. Additionally, if someone has more than one YDS results, the highest one is considered whereas the others are disregarded (Öğrenci Seçme ve Yerleştirme Merkezi, 2017).

3.1.2 Test design and item writing

This section deals with the endorsement of EALTA standards into YDS test design and item writing process, if applied.

3.1.2.1 Do test developers and item writers have relevant experience of teaching at the level of assessment is aimed at?

Item writers are announced to be experts in the field but no more detailed information has been given upon them for security. The experts are recruited either from different universities and schools, albeit solely from Turkey.

3.1.2.2 What training do test developers and item writers have?

It is not quite clear whether the item writers undertake any training before they start writing the test items or not. Yet, they are divided into two groups: one group as the item writers and the other group as the item controllers. Test developers have also their own

language editors both in the foreign language in which the exam will be conducted and the official native language, Turkish.

3.1.2.3 Are there guidelines for test design and item writing?

The item writers are expected to create an archive of questions which are added to the question pool. Therefore, they work full time for writing test items. The item writers are to pursue all resources publicly announced or published in order to prevent any question doublets. The questions are probed in terms of being scientific, comprehensible and appropriate. General item writing principles such as validity, reliability and authenticity are also considered.

3.1.2.4 Are there systematic procedures for review, revision and editing of items and tasks to ensure that they match the test specifications and comply with item writer guidelines?

As above mentioned, there is a systematic procedure followed by test item writers. The groups composed to write and control the test items are the indicators of a peer-review process. Turkish language editors are also at work to evaluate the language used in the test. The problematic test items are opened up for discussion within the group members. If deemed necessary, required amendments are done in compliance with test specifications.

3.1.2.5 What feedback do item writers receive on their work?

OSYM has the question pool of YDS for 30-year-accumulation. Each test item which is chosen to have a place in this pool has to go through numerous steps. At this point, item writing groups together with editorial board evaluate the test items. If approved, the final version gets ready for publication. There is no information reported by OSYM that specific feedback report is given to all test item writers. But the improvements are made on step-by-step basis on the circulation of item design and writing process.

3.1.3 Quality control and test analyses

This section deals with quality control process and procedures for test analyses undertaken for YDS.

3.1.3.1 What quality control procedures are applied?

After the review process on the content, last technical amendments are done. Experts from the fields of foreign language education, language testing and even psychometrics have a sit together for test item writing.

3.1.3.2. Are the tests piloted?

The final version of the prepared tests is not piloted.

3.1.3.3. If there are different versions of the test (e.g. year by year) how is the equivalence verified?

After a serious procedure of preparation, the test bank is molded. OSYM does not seem to publicly mention if there is any system used to enable equivalence among the tests but it renders an equivalence table for YDS and other language proficiency tests conducted.

3.1.3.4 Are markers trained for each test administration?

The scoring is done by machine through automated machines. There is no specific information stated by OSYM that the members of the item writing team are exclusively trained for test administration process.

3.1.3.5. Are benchmarked performances used in the training?

Although there is no overtly announced training session, details on scoring rubrics and marking schemes are given throughout item development process.

3.1.3.6. Is there routine double marking for subjectively marked tests? Is inter and intra-rater reliability calculated?

As automated machines are used for the marking process, there is no subjective marking. Accordingly, double marking makes no sense for it. Henceforth, it seems that no inter and intra-rater reliability is checked during the marking process as the scoring process is enabled automatically.

3.1.3.7 Is the marking routinely monitored?

As there is automated machine scoring, there is no information if someone is assigned for monitoring the marking process. However, there is a double check for the results before they are finally announced to prevent any marking errors.

3.1.3.8 What statistical analyses are used?

As the test is composed of 80 multiple-choice items, multiple-choice option statistics are conducted. Item difficulty and item-total correlation are also reported to be estimated before the release of the results as a rudiment for further tests.

3.1.3.9 What results are reported? How? To whom?

Test takers receive their overall scores from the test conducted. The questions answered correctly and the number of mistakes are also given separately. Additionally, test takers are provided with the rankings as the indicators of their performances from the overall population. The results are announced via an online results system. Each test taker enters the secure login system with a password and user name to see the results. No notification by e-mail is available yet.

3.1.3.10 What process are in place for test takers to make complaints or seek reassessments?

If a test taker is unhappy with the result (s)he gets, (s)he can request his/her test booklet that all the answers are written on. This may take some time but if there is any problem, rescoring may blossom as a need for the compensation of test taker's suffering.

3.1.4 Test administration

This section deals with the test security procedures applied during the administration process.

3.1.4.1 What are the security arrangements?

In order to maintain test integrity, test security is of utmost importance. To ensure maximum test security, OSYM is the big cheese from test development to test implementation phase. The place where the questions are prepared is only accessible to test item writers. The test item writers are to obey the rules of OSYM. For instance, the item writers cannot go out and prepare the items outside and bring them inside. All the questions are to be prepared in the test item preparation room. Other employees other than item writers cannot enter this room as it is forbidden. Moreover, there are security essentials for the publication process, as well. The printing house is a specialized place for the accomplishment of this duty. For this reason, the employees of this printing house cannot have a contact with any other person outside for about 15-20 days. They are to even sleep inside of this printing house. The surrounding area of this printing house is under the supervision of the constabulary, and electronic blackout is activated. Therefore, no cell phone is functional. After the publication of the handouts, the questions are transferred to the exam centers by the courtesy of the police. The place where the questions are kept is also preserved by two guards at the gate. For the implementation phase, the identity and exam entrance cards of each participant are checked by the security before getting into the places where the exam will be held in.

3.1.4.2 Are test administrators trained?

As YDS is a large-scale national examination, very strict administrative procedures are pursued. Therefore, anyone who takes part in the test administration process is to follow the aforementioned strict steps during test administration process.

3.1.4.3 Is the test administration monitored?

Test administration is not observed by any video or audio monitoring system. However, test security is enabled by the invigilators inside and security guards outside. Each test taker with their identity and exam entrance cards is to ensure that they participate by signatures. The test takers are to follow a secured web-application to learn their results by the passwords and usernames given to them on an individual basis.

3.1.4.4. Is there an examiner's report each year or each administration?

As touched upon, the scoring procedures are conducted by automated machines. As mentioned above, there are invigilators who are responsible for each room where the test is applied. In this sense, any problematic situation is reported by them to OSYM via written forms.

3.1.5 Review

In this section, the revision process is touched upon.

3.1.5.1 How often is the test reviewed and revised?

An item bank dating back to 30 years has been developed for YDS. The questions in the item bank are revised even on the preparation phase. Details given by the questions such as numbers, dates etc. have been under a continuous change during the item writing process. Different sets of test items are composed until the final version becomes ready. The revision is implemented by the second group, which is composed of language editors and test item controllers. There is no exact timing procedure on the duration of revisions but the revisions and amendments are made periodically, if necessary.

3.1.5.2 Are validation studies conducted?

Validation phase is triggered with the design of the test items. Item checking and automated scoring involve the functioning of validation studies. This is conducted by experts in the field. However, there is no information if any native speaker is included within the test validation phase as an expert which is a must in many international language examinations.

3.1.5.3 What procedures are in place to ensure that the test keeps pace with changes in the curriculum?

Since YDS does not assess curricular changes albeit the contextual use of foreign languages, it is not based on any curriculum but it pursues the basic tenets of communicative language teaching. In this vein, YDS assesses overall foreign language proficiency instead.

3.1.6 Washback

This section deals with the washback effects of YDS.

3.1.6.1 Is the test intended to initiate change(s) in the curriculum practice?

YDS may seem to initiate change(s) in the current practice as communicative language skills and functional language competences are expected to be included. However, as four language skills are not prominently given place, it may be concluded that real life usages are not frequently emphasized though it is the intended aim within.

3.1.6.2 What is washback effect? What studies have been conducted?

In the literature, the washback effect is described as the influence of testing practices on teaching and learning process (Hughes, 1989; Alderson & Wall, 1993; Bailey, 1996). In essence, the long-term impact of YDS is still not clear since there is a gap on this matter in literature as there are a few studies conducted (Özmen, 2011; Akpınar & Çakıldere, 2013; Akın, 2016; Külekçi, 2016). However, there is a need to label one's foreign language proficiency skills by means of an internationally recognized language test. The nature of YDS tends to exclude listening, speaking and writing skills. Therefore, it may be reported

that language production is not well assessed. Additionally, YDS has some changes within years. A couple of years ago, it was composed of 100 multiple-choice question items, but now it is 80. To add more, the e-YDS has recently mushroomed to cater test takers' needs.

3.1.6.3 Are there preparatory materials?

OSYM provides test takers with sample questions and guidelines for test takers. The questions from the previous years and some technical clues are also given by OSYM to enlighten test takers on its website. Yet, OSYM does not offer any other special preparatory materials for test takers. However, there are many online materials for those who are preparing for YDS though they are not bound to OSYM.

3.1.6.4 Are teachers trained to prepare their students for the text/exam?

Teachers are not exclusively trained to prepare their students for YDS. At the very same, OSYM does not appear to recommend any special teaching resources for them. However, there are many resources available in public.

3.1.7 Linkage to the Common European Framework

This section depicts how YDS is linked to the CEFR, if this is the case.

3.1.7.1. What evidence is there of the quality of process followed to link tests and examinations to the Common European Framework?

OSYM has created a table of equivalence which shows YDS score interpretation and the CEFR levels. However, it may be said that this table is a one-way recognized interpretation as the results of YDS can only be used for local purposes. No item-centered method seems to be applied for the interpretation of the test results. According to the equivalency table proposed by OSYM, A1 refers to somewhere between 30-44 points whereas A2 refers to somewhere between 45-59 points. B1 refers to somewhere between 60-74 points whereas B2 refers to somewhere between 75-94 points. C1 refers to somewhere between 95-99 points whereas C2 refers to 100 points (Öğrenci Seçme ve Yerleştirme Merkezi, 2016b)

3.1.7.2 Have the procedures recommended in the Manual and the Reference Supplement been applied appropriately?

The linking process seems not to be completely enabled as the procedures recommended in the Manual and the Reference Supplement are not applied appropriately. For this reason, it seems reasonable to assume that YDS is not taken valid to be used for international purposes throughout the world.

3.1.7.3 Is there a publicly available report on the linking process?

There is no overtly publicly available report on the linking process, albeit an equivalence table is proposed by OSYM for the interpretation of the results in compliance with the CEFR levels as proposed above.

4 Discussion

YDS is a widely used language proficiency test which is taken by thousands of test candidates. It could be considered as a high stakes test because it has important consequences for the test takers. According to the test scores, test takers' language proficiency is interpreted and evaluated. Nevertheless, as YDS does not seem to assess listening, speaking and writing, test takers are assigned specific scores but those are accounted for performance in vocabulary, grammar and reading, which may mean that all language competences are not represented. Thus, test takers may be in tendency to improve their receptive skills than productive skills (Weiping and Juan, 2005). It could be said the language proficiency is not very likely to be measured with its all dimensions; but, in actual fact, receptive and productive skills are the very fiber of a language.

In addition, as mentioned above in this paper, YDS is used for local purposes. In other words, it is a language proficiency test which is utilized for getting promotions, becoming an academic personnel or pursuing graduate or post-graduate studies within the national borders, mostly. In order to make the language proficiency measurement valid beyond the country, there appears a need for standardization in testing. EALTA guidelines could be employed in order to facilitate good practice in language testing. So, test developers can raise awareness about standard setting for a well-established assessment for proficiency levels. EALTA guidelines offer a broader concept for teachers, teacher trainers and test developers and provide principles regarding respect for the examinees/students, responsibility, fairness, reliability, validity and collaboration (EALTA, 2006). Actually, these guidelines may be considered as a quality control mechanism for test developers in order to check tests' validation, development and fairness. This quality control mechanism is a deep-seated need for language tests as language testing is a big and life changing business because admission to a university, graduation from a university and finding a job can be enabled through passing a language test (Alderson, 2004). Understood in this sense, it may be a central premise for test developers to be meticulous in designing and validating tests.

5 Conclusions

It is probed within the scope of this article whether YDS, as a local but a national language proficiency test, has complied with the EALTA Guidelines. Herein, this study has blossomed two-way alternates. First, YDS has been reviewed in terms of recognition by the internationally set standards. Second, YDS has been reviewed in terms of current applications for the enhancement of a large-scale language test.

In Turkey, foreign language proficiency is aimed to be measured by YDS through 80 multiple-choice test items. In order to prevent misinterpretation of the test scores, OSYM, the official body, points out a table showing scores and their corresponding

references from A to E level, which could be used for local purposes. In actual fact, it may be of crucial importance to have an internationally recognized system which could be used for verification process in developing tests because the same general principles of good assessment practices are applied in planning and developing a test for international test-takers (Young, So & Ockey, 2013).

Furthermore, it is apparent that YDS does not seem to measure listening, speaking and writing separately (Akin, 2016), which may mean that language production is likely to be eliminated. It would be better to include all skills in a language proficiency test in order to measure overall language proficiency. Therefore, a test candidate who takes YDS and gets A still is possibly to have problems in speaking or writing because this A-level may not be a good indicator to project the real language proficiency. Therefore, test takers could sometimes be supposed to take other internationally recognized tests in order to certify their language competences. However, to note herein, translation questions both from Turkish to English, and from English to Turkish are the indicators of a developed ‘transferring skills from and to target language’ (Mirici, 2003), though. Because test takers are expected to trigger both grammatical and textual knowledge in tow in order to comprehend the overall purpose and the underlying meaning in a broader sense.

One more to add, it is a crystal clear fact that it may be suggested more empirical studies be conducted to find out if there is any positive and/or negative washback effect of YDS apart from participants’ commentary. For the generalizability of the test scores and selection of the accurate test items, a piloting stage may be considered with required statistical analyses conducted afterwards. Similarly, test constructs could be thematically arranged pursuant to the representation of the skills so that a standard measurement for different versions of the test could be provided. Also, it may be suggested that new technologies can be taken into account for test administration and implementation processes.

As a result, the EALTA is not the only accurate tool but useful for test development process. YDS within the context of EALTA guidelines might be improved in terms test development process, which has blossomed the need for standardization in language testing and assessment practices in Turkey. This situation directs the path towards the enhancement of the current one in quality. On the other hand, the development of a new internationally recognized language test is also a matter of fact on condition that it is to abide by the principles of accountability, appropriateness and transparency to verify the qualification of the ongoing assessment system in Turkey. This is enabled within the matrix of test developers’ endeavors and participation of the decision-makers into the test development process, so that both good and bad practices can be sorted out.

Acknowledgements

We offer our special thanks to TÜBİTAK for the support and assistance provided.

References

- Akpınar, D. K., & Çakıldere, B. (2013). Washback effects of high-stakes language tests of Turkey (KPDS and UDS) on productive and receptive skills of academic personnel. *Journal of Language and Linguistic Studies*, 9(2), 81-94.
- Alderson, J.C. (2004). *Quality Control in Evaluation and Assessment*. Paper presented at Seminar on Common Standards in Europe, Szeged, April 23rd, 2004.
- Alderson, J. C. (2010). A survey of aviation English tests. *Language Testing*, 27, 51-72.
- Alderson, J.C.&Banerjee, J.(2008). *EALTA's Guidelines for Good Practice: A test of implementation*. Paper presented at the 5th Annual Conference of the European Association for Language Testing and Assessment, Athens, Greece, May. <http://www.ealta.eu.org/conference/2009/program.htm>, last accessed 15th February 2016.
- Alderson, J. C., & Wall, D. (1993). Does Washback Exist? *Applied Linguistics*, 14(2), 115-129.
- Akın, G. (2016). Evaluation of national foreign language test in Turkey. *Asian Journal of Educational Research*, 4(3), 11-21.
- Association of Language Testers in Europe. (1994). *ALTE code of practice*. Association of Language Testers in Europe.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13, 257-279.
- De Jong, J. H.A.L., & Zheng, Y. (2011). *Applying EALTA Guidelines- A practical case study on Pearson Test of English Academic*. Pearson Research Note.
- Demir, D., & Genç, A. (2016). The analysis of foreign language proficiency exam for the state employees based on the CEFR. *Hacettepe University Journal of Education*, 31(1), 53-65.
- European Association for Language Testing and Assessment (2006). *Guidelines for Good Practice in Language Testing and Assessment*.
- Güleç, B. A. (2013). Perceived attributions in foreign language assessment. *International Journal of Language Academy*, 1(1), 1-12.
- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- International Language Testing Association. (2007). *ILTA guidelines for practice*.
- Japanese Language Testing Association. (2002). *JLTA code of good testing practices*.
- Kıray, G. (2015). Macro-structure analysis of reading comprehension paragraphs of KPDS and YDS exams within years 2003-2013. *Hasan Ali Yucel Journal of Education*, 12-1(23), 219-233.
- Külekçi, E. (2016). A concise analysis of the Foreign Language Examination (YDS) in Turkey and its possible washback effects. *International Online Journal of Education and Teaching (IOJET)*, 3(4), 303-315.
- Mirici, İ. H. (2003). The factors affecting the success in English proficiency exams and possible contributions of the internet. *Turkish Online Journal of Distance Education*, 4(1), 1-8.
- Öğrenci Seçme ve Yerleştirme Merkezi (2013a). Retrieved on April 25, 2016 from <http://osym.gov.tr/belge/1-14909/basin-duyurusu-2013-yabanci-dil-bilgisi-seviye-belirlem.html>
- Öğrenci Seçme ve Yerleştirme Merkezi (2013b). Retrieved on April 25, 2016 from <http://www.osym.gov.tr/Eklenti/101.yabanci-dil-esdegerlikleri-250713pdf.pdf?0>
- Öğrenci Seçme ve Yerleştirme Merkezi (2016a). Retrieved on May 14, 2016 from <http://dokuman.osym.gov.tr/pdfdokuman/2016/GENEL/EsdegerlikTablosu26022016.pdf>

- Öğrenci Seçme ve Yerleştirme Merkezi (2016b). Retrieved on May 14, 2016 from <http://dokuman.osym.gov.tr/pdfdokuman/2016/GENEL/EsdegerlikYonergesi26022016.pdf>
- Öğrenci Seçme ve Yerleştirme Merkezi (n.d.-a). Retrieved on April 25, 2016 from <http://www.osym.gov.tr/TR.8860/hakkinda.html>
- Öğrenci Seçme ve Yerleştirme Merkezi (n.d.-b). Retrieved on April 25, 2016 from <http://www.osym.gov.tr/Eklenti/2428,ydsingilizceornekpdf.pdf?0>
- Öğrenci Seçme ve Yerleştirme Merkezi (2017). Retrieved on April 20, 2017 from <http://dokuman.osym.gov.tr/pdfdokuman/2017/YDSILKBAHAR/KILAVUZ09022017.pdf>
- Özmen, K.S. (2011). Washback effects of the inter-university foreign language examination on foreign language competences of candidate academics. *Novatis-ROYAL*, 5(2), 215- 228.
- Weiping, G., & Juan, L. (2005). Test analysis of college students' communicative competence in English. *Asian EFL Journal*, 7 (2), 7.
- Yavuzer, H. & Göver, I. H. (2012). Akademik personelin yabancı dil durumu ve yabancı dil sınavlarına bakışı: Nevşehir örneği. *Nevşehir University Journal of Social Sciences*. 1, 136-158.
- Young, J. W., So, Y., & Ockey, G. J. (2013). *Guidelines for best test development practices to ensure validity and fairness for international English language proficiency assessments*. Educational Testing Service.

ENDNOTES

* This study was partly presented at CLARC2016: Perspectives on Language Planning and Policies Conference which was held in Rijeka/Croatia between the dates of 03-05 June 2016

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the Journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (CC BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).