

Betül Hazal Dinçer, Elena Antonova-Unlu* and Alper Kumcu
Assessing the use of multiple-choice translation items in English proficiency tests: The case of the national English proficiency test in Turkey

<https://doi.org/10.1515/applirev-2019-0064>

Abstract: The use of translation for language teaching and assessment, by and large, has been abandoned with the adoption of audio-lingual and communicative approaches in language teaching. As a result, nowadays translation items are not commonly used for measuring language proficiency in international language proficiency tests (e. g. TOEFL, IELTS).

However, there are several countries that still use translation items in their national language proficiency tests (e. g. Turkey, Japan, China, Romania among others). The present study aims to examine whether or not multiple-choice translation items are an appropriate tool for measuring proficiency in English. To this end, the perceived level of difficulty and validity of multiple-choice translation items in the National English Proficiency Test (YDS) in Turkey were examined. The findings revealed that the participants did significantly better on the translation items than on the rest of the test items. They also perceived the translation items as the easiest among all the rest items in YDS. Moreover, while YDS as a whole indicated a strong validity based on correlation with TOEFL PBT Reading Sample Test, the translation items indicated moderate validity. Importantly, there was a significant difference between the two correlations. These findings suggest that multiple-choice translation items are likely to lower the overall validity of YDS tests, inflate the scores of test-takers and, thus, might be considered as problematic for the quality of the tests.

Keywords: translation items, multiple-choice format, English proficiency tests, construct validity, perceived level of difficulty

*Corresponding author: Elena Antonova-Unlu, Hacettepe University, Beytepe Campus, Ankara 06800, Turkey, E-mail: elenaunlu@gmail.com

Betül Hazal Dinçer, Baskent University, Baglica Campus, Ankara 06810, Turkey, E-mail: hazaldincer@yahoo.com

<https://orcid.org/0000-0002-7730-9327>

Alper Kumcu, Hacettepe University, Beytepe Campus, Ankara 06800, Turkey, E-mail: alperkumcu@gmail.com

1 Introduction

The use of translation items in language proficiency tests dates back to the early years of the former century when first language proficiency tests started to be used (e. g. Certificate of Proficiency in English) (Weir 2005). However, the use of translation was abandoned in time and it does not exist anymore in well-known language proficiency tests such as Test of English as a Foreign Language (TOEFL) and International English Language Testing System (IELTS). Yet, there are several countries that still utilise translation items in their national language tests including but not limited to Japan, China, Romania, Turkey and Taiwan (personal communication with scholars via the Research Gate). Some researchers consider the use of translation items as appropriate for measuring language proficiency (Buck 1992; Sun and Cheng 2013), while others (Kikuchi 2006; Klein-Braley and Smith 1985; Lado 1961) put forward that translation items seem to be inappropriate in language proficiency tests since their use might cause concerns about validity. Though this issue grasped the attention of scholars long ago, still there have been just few studies investigating whether or not translation items are an appropriate tool for measuring language proficiency (Buck 1992; Kikuchi 2006; Klein-Braley 1987; Lado 1961; Stevenson 1985; Sun and Cheng 2013), and no consensus has been reached among the scholars.

To this end, the present study, focusing on the National English Proficiency Test (YDS) in Turkey, examines whether translation items in the English language proficiency test are appropriate for measuring English language proficiency.

The study aims to answer the following research questions:

1. Is the perceived level of difficulty of the translation items in YDS compatible with that of the rest of YDS items?
2. Are the translation items in YDS valid to measure English language proficiency?

The article is structured as follows. First, translation in language teaching and testing will be discussed along with the available research on the use of translation in language testing. Secondly, English language proficiency testing in Turkey will be introduced. Further, the study, methodology and the results will be given. Finally, the findings of the study, limitations and suggestions for future research will be discussed.

1.1 Translation in foreign language teaching and testing

The use of translation in foreign language teaching and testing has been controversial. While translation was the most commonly used tool for language teaching and testing with the grammar-translation method (Brown 2007), it was abandoned with the development of the audio-lingual and communicative approaches (Rivers 1981; Richards and Rodgers 2001), which regard the use of mother tongue in the class as ineffective. For example, the first English language proficiency test, the Certificate of Proficiency in English by Cambridge University (1913), included two translation tasks (from English into French or German, and from French or German into English) (Weir 2005). However, nowadays translation items are not used for measuring language proficiency in well-known and commonly accepted English proficiency tests such as TOEFL, IELTS, Pearson Test of English (PTE), etc. The absence of translation items in the above-mentioned international exams is also related to the fact that learners with different native languages are supposed to take them.

Nevertheless, the use of translation for testing is an on-going practice in certain countries, and Turkey is one of them.

Several researchers have questioned the appropriateness of translation items in language testing and suggested that translation should not be used for this purpose (Klein-Braley 1987; Lado 1961). As reasons for rejecting translation as a testing tool, they stated that translation is a different skill and a good language learner may not necessarily be a good translator. Further, the procedure of scoring when translation items are used may not be objective, and it is not clear what skill translation items actually measure. Klein-Braley and Smith (1985) also put doubts whether or not translation is able to measure language proficiency, and claimed that translation items are likely to lack validity. Along with it, nowadays there is a tendency to define translation as a specific field independent from language teaching and testing, which requires “a firm command of both the source and target language” (Snell-Hornby 1985). Together with studies on ‘the act of translating’ (Catford 1965; Nida 1964; Vinay and Darbelnet 1995), translation has gained its place as a separate discipline, which has its own methods and approaches (Munday 2013).

Following the available tendencies in language teaching and testing as well as those in translation studies, the use of translation in language teaching and testing is being reconsidered. In this respect, Pintado-Gutiérrez (2018) called for interdisciplinarity in language teaching and testing, underlining the difference between the

terms of ‘translation’ and ‘pedagogical’ translation. The latter refers to translating tasks used in foreign language teaching and learning that improve language awareness, accuracy, pragmatic and intercultural competence, creativity, problem solving, autonomy and collaboration, etc. Along with it, the researcher called for further research validating or rejecting the appropriateness of the use of translation items in language teaching and testing.

So far, several studies have contributed to the topic and revealed that using translation in foreign language teaching may be beneficial, especially for improving vocabulary, grammar and reading comprehension skills in the foreign language teaching classroom (Calfoglou 2013; Korošec 2013; Lee 2013 among others). In contrast to the interest in using translation for foreign language teaching, using it in language testing seems to be a neglected area of research, and very limited number of studies have examined the use of translation items in language testing.

1.2 Studies on the use of translation in testing

There have been just a few studies that have investigated the appropriateness of translation items in language testing.

Buck (1992) presented two studies investigating the use of translation as a testing tool. In the first study, production translation items used in Japanese university entrance examinations were investigated in terms of reliability and criterion validity (i. e. correlation of the test scores with a concrete outcome). The data obtained from 121 participants were evaluated by seven raters. The raters used their own judgments for assessing the translation of the participants and gave scores from five to zero. For the validity investigation, a random-deletion cloze test and a multiple-choice comprehension test were used. The findings revealed that the correlations between the raters as well as between the translation items and the tests were strong. It was concluded that “the translations are not measuring anything so very different from more commonly used measures of passage comprehension” (p. 132).

In the second study, Buck (1992) examined the construct validity (i. e. correlation of the test against another valid instrument) of the production translation items and compared them with other commonly used testing methods. Four listening and four reading tests were formed with four methods (short-answer comprehension questions, multiple-choice comprehension questions, gap-filling and translation from the native language into the foreign language and vice versa). The production translation items were found to be as reliable as the other testing methods. Thus, it was argued that translation items in the

production format also have “a good convergent and discriminant validity, suggesting that they have construct validity as tests of listening and reading comprehension” (p. 138).

Kikuchi (2006) aimed to track changes in English proficiency tests in Japan from 1994 to 2004. The difficulty level of the reading passages and the item types were investigated in the study. The findings revealed that though new item types were added in time, most of the items were found to measure the receptive and translation skills. The scholar argued that translation items require certain translation skills apart from proficiency in the language and added “this may lead to problems of validity in that test-taking ability, rather than English proficiency, is being measured” (p. 94).

The most recent study (Sun and Cheng 2013) investigated the appropriateness of translation items in the sentence completion format for assessing second/foreign language proficiency. The validity of the translation items in the College English Test (CET), which is a large scale second/foreign language test in China, was analysed via correlating the participants’ scores obtained from the translation items with those of the other item groups: writing, cloze test, reading comprehension and listening comprehension. The participants’ perception on the translation task was also examined in terms of task demand, task difficulty and the validity of the translation test as a measure of their actual translation ability. The translation items in the sentence completion format indicated moderate validity (< 0.50) with the other item groups, and the questionnaire revealed that the students regarded the difficulty level of the translation tasks as relatively low.

To sum up, the review of the studies examining the use of translation items in testing has revealed that such studies being very limited in number have inconsistent results and leave the question of whether translation items are appropriate for language testing open.

1.3 Language proficiency tests in Turkey

The first national language proficiency test appeared in Turkey in 1982 (<https://www.osym.gov.tr>). Since that time, the test has gone through some modifications and, in 2013 the latest version of it called YDS (*Yabancı Dil Sınavı*, Foreign Language Exam) was released. The aim of YDS is to certify the language proficiency of test-takers, which allows them to get a promotion in their academic and professional careers. The test can be used for applications to post-graduate or graduate studies, all the required language proficiency assessments for state employees, for teacher employment by Ministry of

Education, and all language proficiency requirements by governmental institutions, as it is stated by ÖSYM (2019).

YDS is constructed based on the reading skill with multiple-choice items. It consists of 80 items with item types of vocabulary, grammar, cloze test, sentence completion, translation (three items are translation from English into Turkish, and three items are translation from Turkish into English), reading texts, paragraph completion, paraphrasing, and irrelevant sentence (see Table 1).

Table 1: The item types of YDS.

Type	Number of items
Vocabulary	8
Grammar	8
Cloze Test	10
Sentence Completion	10
Translation	6
Reading Texts	20
Dialogue	5
Paragraph Completion	4
Paraphrasing	4
Irrelevant Sentence	5

The points for the overall test are calculated between 0–100, each question corresponds to 1.25 points. The score levels for the test is as follows: A (90–100), B (80–89), C (70–79), D (60–69), and E (50–59). The time allocated for the test is 180 minutes. It is conducted three times a year as paper-based; and ten times a year as computer-based.

Each of the translation test items is given with five options to choose from, one of which is the correct answer and the other four are distractors (Example 1). The instruction given in the test for the translation items is as follows: “For these questions, choose the most accurate Turkish translation of the sentences in English and the most accurate English translation of the sentences in Turkish” (ÖSYM 2018). The literal translation of the given sentence is generally expected as the correct answer.

Example 1 A multiple-choice translation item (YDS-2018/Spring). Retrieved from: <https://dokuman.osym.gov.tr/pdfdokuman/2018/YDSILKBAHAR/CS/ING24072018.pdf>

38. As members of households established complex relationships with outsiders during the agrarian era, they came under the influence of new rules, structures, and expectations.

- A) Hane fertleri, tarım dönemi boyunca yabancılarla çeşitli ilişkiler kurarak yeni kuralların, yapıların ve beklentilerin etkisi altına girmişlerdir.
- B) Hane fertlerinin yabancılarla karmaşık ilişkiler kurması tarım dönemine denk gelmiştir ve bu dönemde yeni kurallardan, yapılardan ve beklentilerden etkilenmişlerdir.
- C) Tarım döneminde hane fertlerinin yabancılarla çeşitli ilişkiler kurmasının sebebi; yeni kuralların, yapıların ve beklentilerin etkisi altına girmiş olmalarıdır.
- D) Tarım döneminde hane fertleri yabancılarla karmaşık ilişkiler kurdukça yeni kuralları, yapıları ve beklentilerin etkisi altına girmişlerdir.
- E) Yeni kuralların, yapıların ve beklentilerin etkisi altına giren hane fertleri, tarım döneminde yabancılarla daha karmaşık ilişkiler kurmuşlardır.

2 Methodology

2.1 Participants

32 healthy participants with normal or corrected-to-normal vision took part in the study (20 females, $M_{\text{age}} = 24.6$ $SD = 3.2$, range = 18–31). Convenience and volunteer sampling were adopted considering the duration of the testing procedure. All the participants were native speakers of Turkish who were learners of English as a foreign/second language (based on self-reports). 28 of the participants (87%) reported to have taken one of YDSs before. Detailed information about the materials and the test procedure was given to all the participants, and they gave written consent for participating in the study. The study was covered by ethics approval from Hacettepe University, Turkey.

2.2 Materials

YDS Spring 2018 and TOEFL Paper-Based Test (PBT) Reading Sample Test were given the participants (hereinafter referred to as TOEFL). YDS Spring 2018 was the latest YDS published online by ÖSYM on their official website as of the current study. TOEFL was decided to be used as the benchmark for the current study as YDS is constructed based on reading skill exclusively.

A questionnaire collecting information about the participants' background and their perception of the level of difficulty regarding the test times was also administered. The participants were requested to assess the level of difficulty of each section of the items in the YDS(s) that they had taken before based on a Likert scale from 1 to 5 (1 – very easy, 5 – very difficult).

2.3 Procedure

The participants attended separate sessions for each test with an interval of a week. The questionnaire and YDS were given in the first session. TOEFL was given in the second one. Paper-based tests were used in the study. A study room at the Hacettepe University Library and various classes at the university were used to present an authentic test environment. The sessions were arranged according to the availability of the participants, with one to four participants per session. Thus, ten sessions were conducted for collecting the data on different days. Participants were allocated the same duration allocated for the real YDS and TOEFL (180 minutes for YDS and 60 minutes for TOEFL). Dictionaries were not allowed in neither of the tests. The two tests took four hours to complete for each participant.

2.4 Data analysis

The data were analysed and visualised with R programming language version 3.6.0 (R Core Team 2019). Descriptive data were analysed and presented with *psych* package (Revelle 2018). Linear mixed-effect modelling with *lme4* package (Bates et al. 2015) and t-test were used to analyse performance of the participants in the translation items as compared to the other item types in YDS.

To reinforce the performance results, the data from the difficulty perception questionnaire were analysed by attaining ordinal scores to each difficulty level (i. e. very easy = 1, easy = 2, medium = 3, difficult = 4, very difficult = 5). Then, the difficulty scores corresponding to translation items were contrasted against the scores for the other sections averaged into one section.

Construct validity of YDS as a whole and of the translation items was examined based on Pearson's correlation coefficients to reveal whether translation items in YDS are valid to measure English language proficiency. Construct validity refers to how well a test measures what it supposes to measure and is traditionally measured via correlation of the test against another valid instrument (Lado 1961). A conventional approach was chosen to interpret correlation

coefficients (Schober and Schwarte 2018). Correlations were compared based on Fisher normalisation with *cocor* package (Diedenhofen and Musch 2015).

3 Results

3.1 Performance and perception of difficulty

Overall scores in YDS and TOEFL (out of 100) and the scores in each item section of YDS (out of 100) collected from the 32 participants were calculated. Table 2 presents the descriptive statistics of the data.

Table 2: Descriptive statistics of YDS and TOEFL scores collected from 32 participants.

Test	Mean	SD	Min.	Max.	Skewness	Kurtosis	SE
TOEFL	53.47	23.46	15	93	-0.07	-1.26	4.15
YDS	65.39	23.90	22.50	98.75	-0.29	-1.18	4.23
Item type							
Translation	76.38	26.17	16	100	-0.65	-0.95	4.63
Cloze test	73.12	24.68	20	100	-0.44	-1.02	4.36
Dialogue	70.62	27.82	0	100	-0.69	-0.44	4.92
Sentence completion	70.00	24.89	20	100	-0.36	-1.04	4.4
Irrelevant sentence	67.50	29.95	0	100	-0.74	-0.52	5.29
Vocabulary	66.80	29.89	0	100	-0.48	-0.84	5.28
Grammar	66.02	27.54	0	100	-0.49	-0.77	4.87
Reading texts	59.06	25.03	15	100	0.01	-1.25	4.42
Paragraph completion	54.69	32.03	0	100	-0.25	-1.12	5.66
Paraphrasing	53.91	31.18	0	100	-0.19	-0.97	5.51

A linear mixed-effects model with the participants as random effect indicated that the participants performed significantly better in the translation items compared to paraphrasing ($\beta = 22.47$, $t = 5.80$, $p < 0.0001$), paragraph completion ($\beta = 21.69$, $t = 5.60$, $p < 0.0001$), reading texts ($\beta = 17.31$, $t = 4.47$, $p < 0.0001$), grammar ($\beta = 10.36$, $t = 2.67$, $p = 0.008$), vocabulary ($\beta = 9.58$, $t = 2.47$, $p = 0.01$) and irrelevant sentence items ($\beta = 8.87$, $t = 2.29$, $p = 0.02$) (see Figure 1). As a result, the participants overall performed better in YDS ($Mean = 65.81$, $SD = 23.87$) against YDS minus translation items (i.e. the rest of the items) ($Mean = 64.64$, $SD = 24.24$); $t(31) = 3.79$, $p = 0.0006$.

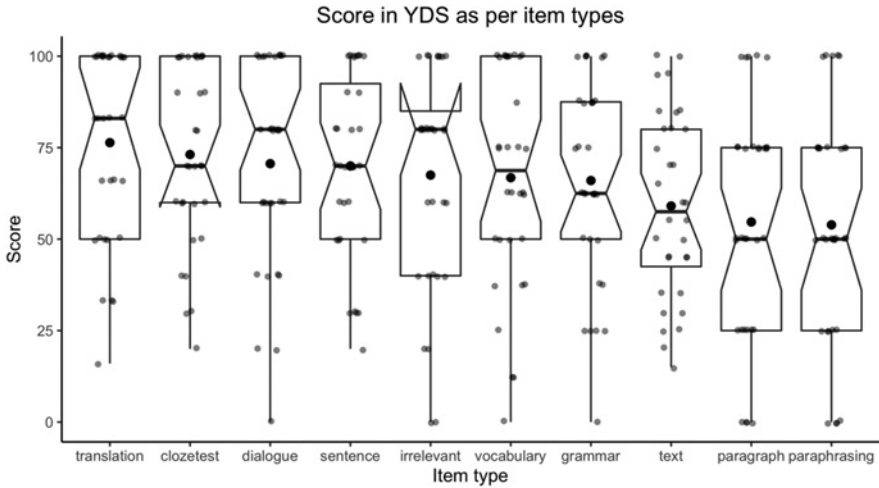


Figure 1: Mean and raw scores across item types in YDS in the descending order. Notched box plots show median (horizontal line), mean (black dot), 95% confidence interval of the median (notch), interquartile range (the box), the first and the third quartiles (lower and upper ends of the box) and ranges (vertical line). Grey dots represent data points.

Further, the results of the difficulty perception questionnaire indicated that both types of the translation items were found significantly easier (translation from English to Turkish: $t(25) = 5.59$, $p < 0.0001$; translation from Turkish to English: $t(25) = 5.70$, $p < 0.0001$) as compared to the mean ranking of the rest of the items averaged into one section.

3.2 Validity

Pearson's correlation indicated that there is a positive, strong correlation between overall scores in YDS and overall scores in TOEFL; $r(30) = 0.88$, $p < 0.0001$, indicating that YDS, in general, can be interpreted as a valid test for measuring English language proficiency based on the reading skill. On the other hand, there was a positive but medium correlation between the score in translation items in YDS and the overall scores in TOEFL; $r(30) = 0.61$, $p = 0.0002$. We transformed both correlation coefficients into z scores with Fisher normalisation to be able to compare them. The results demonstrated that the validity of translation items based on correlation coefficients is significantly lower than that of the overall scores in YDS ($z = -3.96$, $p = 0.0001$) (see Figure 2).

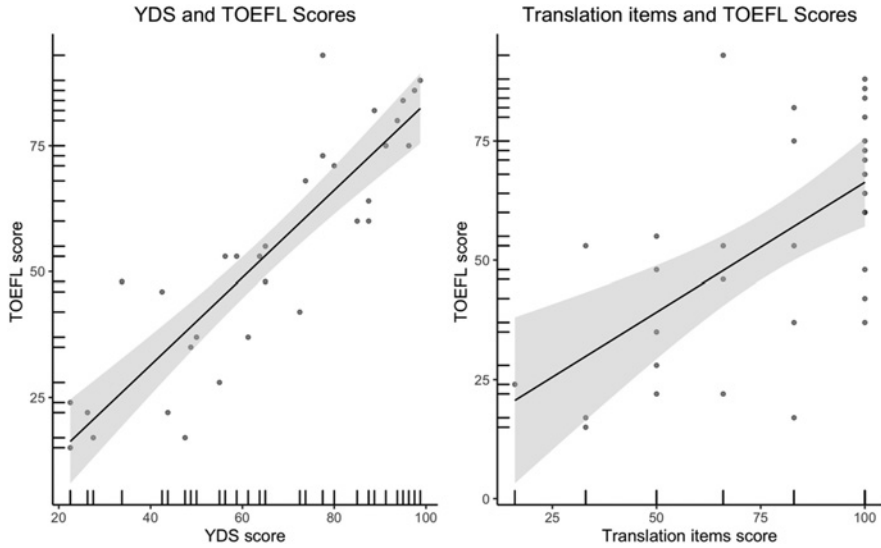


Figure 2: Scatterplots showing the correlations between YDS score and TOEFL score and translation items score and TOEFL score respectively. Scatterplots have linear regression lines. Blue bands around the lines represents 95% confidence interval. Tassels at the x and y axis illustrate the marginal distribution of data along x and y variables.

4 Discussion

The present study aimed to examine whether or not multiple-choice translation items are an appropriate tool for measuring proficiency in English. For this purpose, first, the level of difficulty of multiple-choice translation items based on the participants' performance in YDS as well as the participants' perception of multiple-choice translation item difficulty in YDSs they had taken before was examined. Second, the validity of YDS and of multiple-choice translation items in YDS was checked via correlating the scores of the participants obtained from YDS with those obtained from TOEFL.

According to the results, the participants did significantly better on the translation items when compared with the rest of the test items. The perception analyses obtained via the questionnaire revealed that they also considered multiple-choice translation items as the easiest item type among the all. This finding suggests that translation items in the multiple choice format might fail to reveal the actual level of the test-takers' English proficiency and that the use of multiple-choice translation items is likely to inflate participants scores.

Further, YDS indicated a positive, strong validity (0.88) with TOEFL, which means that YDS can be interpreted as a valid language proficiency test for the reading skill. Then, the scores of the participants obtained from the translation items were correlated with those of TOEFL. The multiple-choice translation items were found to have a positive, moderate validity (0.61) with TOEFL. Fisherman normalisation was used to reveal whether there is a significant difference between the two correlations. The results demonstrated that there is a significant difference between the two correlation coefficients. This finding may suggest that translation items in the multiple choice format lower the overall validity of the test and, thus, its quality.

In this respect, our findings are consistent with the results of Sun and Cheng (2013), who also found that translation items in the sentence completion format have a moderate validity. Moderate validity does not exclude the use of translation items for assessing English proficiency. However, taking into consideration that the overall validity of YDS was significantly higher than that of the translation items, it seems possible to put forward that the scores obtained for multiple-choice translation items seem not to reflect the proficiency level in English of test-takers in the most objective way.

Concerns about the validity of translation tasks have been around for many years as the literature review above showed (see Buck 1992; Kikuchi 2006; Klein-Braley and Smith 1985). The validity of translation items in the production format was questioned because of the scoring procedure, which was argued to be subjective and problematic in terms of fairness. Translation items in the multiple-choice format are easier to score and are considered to provide an objective scoring, however, change in format from production to multiple-choice is unlikely to solve the problem and eliminate the validity issue.

Considering the wide extent to which the scores of national proficiency tests are utilised and the impact the scores may have in people's lives, the validity evidence should be provided for each item types, and all the items should be formed very carefully. So far, several studies have revealed that using translation may be useful in teaching vocabulary, grammar knowledge and reading comprehension (Calfoglou 2013; Korošec 2013; Lee 2013). However, for assessing these skills, translation items seem to be not the best option. For this reason, other item types such as filling in the gaps, paraphrasing, cloze tests, matching, may be more appropriate, as suggested by several scholars (e. g. Aitken 1975; Hughes 1989).

Last but not least, the use of translation items in language proficiency test may also cause a misconception among test-takers: they may transfer the level of their performance on translation items to their general ability to translate. Considering the lowest difficulty level of multiple-choice translation items

among all the test items, most of the test-takers are likely to consider themselves as good translators. This tendency may be also harmful for the field of translation studies and translation as a profession.

To finalize the discussion, it is necessary to point out the limitations of the present research and to put forward some suggestions that could be addressed in future studies. First, the present study was limited to the data collected from 32 participants. Though the participants were of various educational backgrounds and had various levels of English proficiency, increasing the number of participants would contribute to the research and provide statistically more accurate results. Second, to find out whether multiple-choice translation items are valid to measure English language proficiency construct validity based on Pearson's correlation coefficients was examined in this study. Further research should investigate the content validity, criterion-based validity and/or face validity of translation items as well as conduct item analysis to get a deeper insight into the issue. Third, the present study focused on the translation items used in the latest version of YDS. Including data from several versions of YDS would add to the findings of the present study. Additionally, more research on the use of multiple-choice translation items as well as translation items of different formats (completion, production, etc.) used in national proficiency exams in other countries will help understand whether or not translation items are an appropriate tool for assessing English language proficiency.

5 Conclusion

The present study demonstrated that the use of translation items in the multiple-choice format for testing language proficiency lowers the validity of the test and thus, its quality. The level of difficulty of translation multiple-choice items was also found incompatible when compared with the rest of the test items. The findings of the present study suggest that the use of translation items in language proficiency tests should be reconsidered.

References

- Aitken, Kenneth G. 1975. Problems in cloze testing re-examined. *TESL Reporter* 8(2). 9–16.
- Bates, Douglas M, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using Lme4. *Journal of Statistical Software* 67(1). 1–48. <https://doi.org/10.18637/jss.v067.i01>.

- Brown, H. Douglas. 2007. *Principles of language learning and teaching*. White Plains: Pearson Longman.
- Buck, Gary. 1992. Translation as a language testing procedure: Does it work? *Language Testing* 9(2). 123–148. <https://doi.org/10.1177/026553229200900202>.
- Calfoglou, Christine. 2013. An optimality translation proposal for the foreign language class. In Dina Tsagari & Georgios Floros (eds.), *Translation in language teaching and assessment*, 93–114. Newcastle upon Tyne, UK: Cambridge Scholars Publishing.
- Carreres, Angeles. 2006. Strange bedfellows: Translation and language teaching: The teaching of translation into L2 in modern languages degrees; Uses and limitations. *Sixth Symposium on Translation, Terminology and Interpretation in Cuba and Canada. Canadian Translators, Terminologists and Interpreters Council*, 1–21.
- Catford, John Cunnison. 1965. *A linguistic theory of translation: An essay in applied linguistics*. Oxford: Oxford University Press.
- Diedenhofen, B. & J. Musch. 2015. Cocor: A comprehensive solution for the statistical comparison of correlation. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0131499>.
- Hughes, Arthur. 1989. *Testing for language teachers*. Cambridge: Cambridge University Press.
- Kikuchi, Keita. 2006. Revisiting English entrance examinations at Japanese universities after a decade. *JALT Journal* 28(1). 77–96.
- Klein-Braley, Christine. 1987. Fossil at large: Translation as a language testing procedure. In R. Grotjahn, C. Klein-Braley & D.K. Stevenson (eds.), *Taking their measure: The validity and validation of language tests*, 111–132. Bochum: Brockmeyer.
- Klein-Braley, Christine & Veronica Smith. 1985. Incalculable and full of risks? Translation L1 to L2 as a testing procedure. In Christopher Titford & A. E. Hiege (eds.), *Translation in foreign language teaching and testing*, 155–168. Tübingen: Gunter Narr Verlag Tübingen.
- Korošec, Melita Koletnik. 2013. Teaching grammar through translation. In Dina Tsagari & Georgios Floros (eds.), *Translation in language teaching and assessment*, 23–40. Newcastle upon Tyne, UK: Cambridge Scholars Publishing.
- Lado, Robert. 1961. *Language testing: The construction and use of foreign language tests: A teacher's book*. Longmans: London.
- Lee, Tzu-yi. 2013. Incorporating translation into the language classroom and its potential impacts upon L2 learners. In Dina Tsagari & Georgios Floros (eds.), *Translation in language teaching and assessment*, 3–22. Newcastle upon Tyne, UK: Cambridge Scholars Publishing.
- Munday, Jeremy. 2013. *Introducing translation studies: Theories and applications*, 2nd edn. Amsterdam: Routledge.
- Nida, Eugene A. 1964. *Toward a science of translating: With special reference to principles and procedures involved in Bible translating*. Leiden: E. J. Brill.
- ÖSYM. 2019. *Yabancı Dil Seviye Tespit Sınavı (YDS) Kılavuzu*. <https://dokuman.osym.gov.tr/pdfdokuman/2019/YDS-1/kilavuz30012019.pdf> (accessed 1 April 2019)
- ÖSYM. *Yabancı Dil Seviye Tespit Sınavı (YDS) İngilizce İlkbahar Dönemi 1 Nisan 2018*. <https://dokuman.osym.gov.tr/pdfdokuman/2018/YDSILKBAHAR/CS/ING24072018.pdf> (accessed 1 November 2018).
- Pintado-Gutiérrez, Lucía. 2018. Translation in language teaching, pedagogical translation, and code-switching: Restructuring the boundaries. *Language Learning Journal* 1–21. <https://doi.org/10.1080/09571736.2018.1534260>.
- R Core Team. 2019. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.r-project.org/>.

- Revelle, William. 2018. Psych: Procedures for personality and psychological research. Evanston, Illinois, USA: Northwestern University. <https://cran.r-project.org/package=psych>.
- Richards, Jack C. & Theodore S. Rodgers. 2001. *Approaches and methods in language teaching*. Cambridge: Cambridge University Press.
- Rivers, Wilga M. 1981. *Teaching foreign-language skills*. Chicago: University of Chicago.
- Schober, Patrick & Lothar A. Schwarte. 2018. Correlation coefficients: Appropriate use and interpretation. *Anesthesia and Analgesia* 126(5). 1763–1768. <https://doi.org/10.1213/ANE.0000000000002864>.
- Snell-Hornby, Mary. 1985. Translation as a means of integrating language teaching and linguistics. In Christopher Titford & A. E. Hieke (eds.), *Translation in foreign language teaching and testing*, 21–28. Tübingen: Gunter Narr Verlag Tübingen.
- Stevenson, Douglas K. 1985. In other words: Language testers and translation tests. In Christopher Titford & A. E. Hieke (eds.), *Translation in foreign language teaching and testing*, 137–153. Tübingen: Gunter Narr Verlag Tübingen.
- Sun, Youyi & Liying Cheng. 2013. Assessing second/foreign language competence using translation: The case of the college English test in China. In Dina Tsagari & Georgios Floros (eds.), *Translation in language teaching and assessment*, 235–252. Newcastle upon Tyne, UK: Cambridge Scholars Publishing.
- Vinay, Jean-Paul & Jean Darbelnet. 1995. *Comparative stylistics of French and English: A methodology for translation*. Translated by Juan C. Sager and Marie-Josée Hamel. Amsterdam and Philadelphia: John Benjamins Publishing.
- Weir, Cyril J. 2005. *Language testing and validation*. Palgrave Macmillan UK: Basingstoke.

Article note: This paper is a revised version of an unpublished MA thesis written by B. Dinçer and supervised by E. Antonova-Unlu.