SN Social Sciences (2024) 4:151
https://doi.org/10.1007/s43545-024-00930-9

ORIGINAL PAPER

Development of the Turkish Author Recognition Task (TART) and the Turkish Vocabulary Size Test (TurVoST)

Tan Arda Gedik[1,2]

Received: 2 November 2023 / Accepted: 19 June 2024 / Published online: 12 August 2024

Abstract

This article reports the development of two novel research tools for Turkish, the Turkish Author Recognition Task (TART) and the Turkish Vocabulary Size Test (TurVoST). Such tools have been readily available for English, Spanish, Korean, Dutch and Chinese but not for Turkish. These tools help researchers to identify the print exposure levels of L1 speakers and an approximation of L1 speakers' receptive vocabulary knowledge, respectively. Measuring print exposure is important as it is an important driver of L1 development from a usage-based perspective (e.g., Dąbrowska in Cognition 178:222–235, 2018), which influences vocabulary, grammar, and collocation knowledge. The findings show that the TART and TurVoST are significantly correlated at 0.47 and the TART accounts for almost 18% of the variance in vocabulary knowledge. Reliability (Cronbach's alpha) scores were found to be 0.99 and 0.74 for two tests respectively. In light of similar previous studies of various ARTs and vocabulary size tests, the TART and the TurVoST are found to be reliable research instruments with correlations and reliability scores within the range of what has been reported in the literature. Potential uses of these two instruments are discussed. All data, R codes, and research instruments are publicly available at https://osf.io/u6t8m/?view_only=63cf706c381a4214950984dae5470df6.

Keywords Turkish · Author recognition task · Vocabulary size test · Development · Usage-based approaches · Literacy effects

✉ Tan Arda Gedik
tan.gedik@fau.de

[1] FAU Erlangen-Nürnberg, Language and Cognition Lab, Erlangen, Germany

[2] Department of Psychology, Bilkent University, Ankara, Turkey

Vocabulary knowledge and reading

Exposure to language is vital from a usage-based perspective because variances in exposure have substantial outcomes in how speakers' linguistic performance develops (e.g., Divjak 2019; Goldberg 2006, 2019; Dąbrowska 2015). First language (L1) speakers show many individual differences (IDs) in their linguistic knowledge because of differences in how much print exposure they receive (e.g., Dąbrowska 2018; Kidd et al. 2018). Such IDs are important for linguistic theories because so far, the received conventional wisdom has been that L1 speakers converge on the same linguistic knowledge uniformly (see Dąbrowska 2015)[1]. Recent studies show that this may not be the case, and that speakers of different languages show print exposure related IDs (see Kidd et al. 2018; Dąbrowska 2020), because language learning is a usage-based phenomenon.

Vocabulary knowledge may exhibit the most IDs as a result of experience with written language when compared to other types of linguistic knowledge. Dąbrowska (2018) demonstrates that across vocabulary, grammar, and collocational knowledge, it is vocabulary that correlates the strongest with print exposure, measured by an author recognition task. Reading enhances the vocabulary knowledge of speakers positively (e.g., Cunningham and Stanovich 1998; Burt and Fury 2000; Grant et al. 2007; Stanovich and Cunningham 1992; Stanovich et al. 1995; West and Stanovich 1991). This is unsurprising as written language covers more varied language when compared to spoken language (e.g., Cunningham and Stanovich 1992; Hayes and Ahrens 1988). Table 1 showcases this difference.

Hayes and Ahrens (1988) show that written language, even children's books, are lexically more enriched than spoken language output of highly educated speakers, as measured by the number of rare words (defined as the words with a frequency of 10,000 or more in the reference word list). Cunningham and Stanovich (1998) analyze the lexical differences between spoken and written child and adult texts. Their detailed summary reveals that written texts are lexically more enriched than spoken texts, resembling the findings of Hayes and Ahrens (1988). As a result of this, non-basic vocabulary is thought to be acquired later in adulthood incidentally through exposure to written materials (see Dąbrowska 2009 for a more detailed discussion).

Exposure to written materials has an advantage in vocabulary knowledge. Researchers have established statistically significant correlations varying anywhere between 0.40 and 0.80 between vocabulary knowledge and print exposure (e.g., Cunningham and Stanovich 1998; Dąbrowska 2018; De la Garza, in preparation; Stanovich and Cunningham 1992; see also Table 5 in this paper). Correlations remain significant and at the same value even after controlling for other variables such as reading comprehension and nonverbal IQ skills in the above-mentioned studies. This suggests that higher abilities such as word-meaning inferencing are not responsible for above average vocabulary knowledge, rather, there is a direct correlation between reading and vocabulary knowledge of a speaker.

[1] Generative approaches (especially the minimalist program) accounts for high variability across speakers in vocabulary knowledge by locating lexis in the periphery and arguing that it is subject to factors beyond the universal grammar (e.g., Chomsky 1965; Eisenbeiß 2009).

Table 1 Richness of vocabulary across selected written and spoken modalities (adapted from Hayes and Ahrens 1988)

	Proportion of text from 5000 basic lexicon	Rank of median Word	Number of rare words per 1000 tokens
College graduates in conversation with friends and spouses	0.94	496	17.3
Popular prime time tv	0.94	490	22.7
Children's books	0.92	627	30.9
Adult books	0.88	1058	52.7
Newspapers	0.84	1690	68.3
Scientific articles	0.70	4389	128.0

Multiple studies provide comprehensive insights into the correlation between reading habits and vocabulary acquisition. Dąbrowska (2018) conducted research to explore the impact of literacy on linguistic proficiency in adult native English speakers. Her findings revealed a statistically significant relationship between exposure to printed material and vocabulary knowledge. Notably, Dąbrowska identified print exposure as the primary determinant of vocabulary knowledge, accounting for approximately 25.8% of the variability observed. Moreover, she highlighted the combined influence of education and print exposure as another significant predictor. According to Dąbrowska, individuals with limited schooling and minimal exposure to printed material face the greatest challenges in vocabulary knowledge. Conversely, those with extensive education and substantial print exposure exhibited the highest levels of vocabulary knowledge.

As we can see in this brief overview, vocabulary knowledge and reading appear to be in a relationship. While there are author recognition tasks designed for English (Acheson et al. 2008), Spanish (De la Garza, in preparation), Chinese (Chen and Fang 2015), Korean (Lee et al. 2019), and Dutch (Brysbaert et al. 2020), such a task for Turkish has not yet been developed. Similarly, although vocabulary size tests were developed for various other languages, Turkish does not yet have a vocabulary size test. Therefore, to bridge the research gap, this paper reports the development of the Turkish Author Recognition Task (TART), and the Turkish Vocabulary Size Test (TurVoST).

Author recognition tasks

Author recognition tasks (ARTs) are widely used as a proxy measure for measuring how much a speaker has cumulatively read in their life. It is more valid than a questionnaire collecting information on print exposure, because people tend to give socially desirable answers on print exposure questionnaires (e.g., Acheson et al. 2008). Usually, ARTs present a list of real and fake author names to the participants, asking participants to mark the familiar names.

ARTs are known to correlate well with reading comprehension, spelling skills, and up to a certain extent with nonverbal IQ (Dąbrowska 2018; Mar and Rain 2015; Payne et al. 2012; Stanovich and West 1989). They also correlate well with different

interfaces of linguistic knowledge, such as grammar and collocations, as mentioned earlier. This is quite important from a usage-based perspective because ARTs prove to be a reliable predictor of print exposure, a phenomenon known to affect linguistic knowledge as written language provides more enriched language to the speaker. ARTs provide a fast and reliable way of providing information for speakers' print exposure.

Various ARTs have been developed for other languages. Acheson and colleagues' (2008) version of the English ART has been the most frequently used ART for studies analyzing English speakers. It consists of 65 real and 65 foil names. Several studies using the ART show that it can predict vocabulary knowledge. Dąbrowska (2018) found a strong correlation between the English ART and a modified version of the English Vocabulary Size Test (VST) ($r=0.60$). Similarly, James and colleagues (2018) found a correlation of 0.45 between the scores on the English ART and the English VST.

Brysbaert and colleagues (2020) developed the Dutch ART to be used in the Netherlands and Belgium. They first sampled 15,000 fiction authors from the library of Ghent, combined it with 7500 foil names and piloted this with a total of 25,000 Dutch speakers across the two countries. Each participant received 70 random real author names and 30 random foil names. In the end, they picked the top 90 real and 42 foil author names that were known to at least half of the participants or more. Their test had a split-reliability score of 0.90 and the test correlated at 0.42 with the Dutch VST. Similar to how Brysbaert and colleagues operationalized the Dutch ART, Lee et al. (2019) developed the Korean ART, consisting of 40 popular and 40 foil names. Their study showed higher correlations between the ART and vocabulary knowledge, reading comprehension, and the accuracy in a lexical decision task than self-reported reading ($r=0.35$, $r=0.31$; $r=0.39$, correlations between the variables and the Korean ART, respectively).

De la Garza (in preparation) developed the Spanish ART, with 81 real and 41 foil names. She first sampled 200 popular Spanish author names from websites, magazines, and national libraries in Mexico. Then, she piloted the 200 names with 150 Spanish speakers in Mexico and took the top 81 real names. The Spanish ART correlated with a Spanish vocabulary test significantly ($r=0.37$) and with a nonce-verb inflection task ($r=0.19$). Her study also calculated delta prime scores to counterbalance the binary nature of the ART. The Spanish ART correlated more strongly with delta-prime scores (vocabulary $r=0.44$; nonce-word inflection $r=0.25$).

Currently, there exists no author recognition tests for Turkish. This creates a dilemma for usage-based linguists working with L1 Turkish speakers: there is no unified and economic way of measuring the effects of reading in Turkish on L1 Turkish native speakers and their linguistic knowledge (as ARTs are employed as such in other studies, see Dąbrowska 2015, 2018).

In addition to this, the Turkish Author Recognition Test (TART) can be useful in several other ways. For instance, in applied linguistics, the TART can serve as a valuable tool for evaluating collective exposure to written language and tracking language development over time for native Turkish speakers, and if normed, it can also serve a purpose to track L2 speakers' exposure to written Turkish, too. Moreover, in education, the TART can inform curriculum development by identifying areas where

learners may need additional support in reading comprehension if reading comprehension tasks are used in combination with it. Finally, a norming study of the TART with a larger sample of native Turkish speakers may show how much the general population does reading and how this differs among different educational or socio-economic backgrounds.

Such a study would have interesting findings if compared against the results of other ARTs, indicating collective reading trends of nations. Having the TART can allow linguists to work with bilingual or heritage speakers whose other languages have ARTs, and this would provide a more equal ground to measure experience to written language in both languages—especially in the case of heritage speakers, whereby such speakers receive little to no literacy training or formal schooling for/in their heritage language. The findings from such applications can prove useful in cross-cultural or cross-linguistic situations to track general trends across populations.

Vocabulary size tests

VSTs are a fast and economical way of testing written receptive vocabulary using a variety of formats. As mentioned earlier, VSTs correlate strongly with print exposure measures (e.g., Cunningham and Stanovich 1998), spelling (e.g., Stanovich and West 1989), reading comprehension (West and Stanovich 1991), and also with other linguistic knowledge such as collocations and grammar (e.g., Dąbrowska 2018).

Although VSTs have been a popular research instrument, especially to measure L2 speakers' vocabulary knowledge, there is also some criticism. Critics of frequency based VSTs argue that such assessments may offer limited insights into individuals' true vocabulary size. These critiques (Gyllstad et al. 2021; Stewart et al. 2021; Stoeckel et al. 2021) highlight several key concerns. First, they suggest that VSTs focused solely on high-frequency words may fail to capture the full extent of individuals' vocabulary knowledge. Second, critics argue that these tests lack sensitivity in detecting vocabulary knowledge, as emphasized by Stewart and colleagues (2021). Finally, Stoeckel et al. (2021) discuss that word frequency is not the one and only driving factor behind vocabulary acquisition or accuracy on these vocabulary items.

While these are rightful concerns, these do not apply to the current test at hand as it is developed for L1 Turkish speakers. Some counterarguments drawn from studies such as Schmitt and Schmitt (2014) suggest that frequency based VSTs still offer valuable insights into vocabulary acquisition and proficiency. Similarly, Stoeckel and colleagues (2021) also mention that the current VST tests remain useful for purposes despite their criticisms. These studies (e.g., Schmitt and Schmitt 2014) demonstrate that while frequency-based assessments may not provide a comprehensive measure of vocabulary knowledge, they remain useful tools for gauging general vocabulary knowledge and tracking language development over time, at least for L2 speakers. What these discussions show is that one should be careful with the vocabulary size estimations based on VSTs– and this rightfully applies to L1 speakers as well, especially with the criticism that not all low-frequency items from a frequency band (e.g., 14 K) are at the same level of difficulty. Thus, while it is tempting, we suggest avoiding making vocabulary size estimations using the TurVoST or other VSTs.

Nevertheless, VSTs still remain a valid test to track general vocabulary development, and also continue to play a valuable role in linguistic knowledge assessment when used in conjunction with other measures or linguistic tasks in L1 acquisition studies, which have important psycholinguistic implications as well as implications for general linguistic theory. For instance, Dąbrowska (2018) shows that L1 speakers with more vocabulary knowledge also have a better grasp of collocations and grammar to a smaller extent, as vocabulary co-occurs with larger structures and speakers track these co-occurrences when constructing their L1. This provides further evidence to the idea that vocabulary knowledge is tied together with other larger linguistic units.

However, to this day, there exists no publicly available Turkish VSTs or measures that can be administered to adults or studies that have examined the vocabulary knowledge of L1 Turkish speakers. There exists one vocabulary test for Turkish children that aims to measure both receptive and productive vocabulary (Berument and Güven 2013), and an adapted version of the Peabody Picture Vocabulary Test (Katz et al. 1974), again to be used with children. Therefore, constructing the TurVoST is of great importance for cross-linguistic comparisons to investigate the effects of print exposure on receptive vocabulary knowledge among adult L1 Turkish speakers.

The TurVoST tests receptive vocabulary in Turkish, resembling the original VST for English (Nation and Beglar 2007) and utilizes the multiple-choice question format. Constructing the TurVoST was conducted in line with Nation and Beglar (2007) (available at http://www.victoria.ac.nz/lals/paul-nation/nation.aspx), as it has been a popular and reliable VST used in many English linguistics studies (e.g., Beglar 2010; McLean et al. 2014).

The main use case of the TurVoST is in combination with other linguistic tasks that are administered to L1 Turkish speakers. The findings between such tasks and the TurVoST can be used determine the effect of vocabulary knowledge on morphological productivity (as in Dąbrowska, 2008) or the relationship between The TurVoST can be used in a variety of contexts. Firstly, in L1 education in K-12, the results of the TurVoST can provide educators with some insight into vocabulary development. Secondly, when the TurVoST is used in combination with the TART, it can provide opportunities for cross-linguistic investigations. This would be include both Turkey Turkish-Cypriot Turkish speakers as well as L1 Turkish speakers and L1 English speakers (among other combinations). Furthermore, the vocabulary knowledge and its relation to reading can be examined among functional illiterates (i.e., adult speakers who had very little formal education), low academic attainment (i.e., adult speakers who did not receive any university education), and high academic attainment speakers (i.e., adult speakers who finished an undergraduate degree and may have pursued further degrees) across other languages where similar tests also exist. As such, this would show if literacy, education and socioeconomic status that is tied to education affect vocabulary knowledge among L1 Turkish speakers. Again, when normed, the TurVoST may prove useful for L2 Turkish studies to determine vocabulary knowledge, although in that case additional precautions should be taken (see the discussion in Stoeckel et al. 2021). Finally, the TurVoST can be used among heritage speakers of Turkish to measure their vocabulary knowledge to use in combination

with other linguistic tasks. This would provide a basis to compare L1 and heritage Turkish speakers' vocabulary knowledge.

Methodology

The Turkish Corpus (TSCorpus) (Sezer and Sezer 2013) is a web-based corpus with 491,360,398 words. The TSCorpus can create a lemmatized frequency list of the entire corpus and this can be downloaded in.txt format. This file provides the ranking of the word in the whole corpus as well as the raw frequency. We downloaded this file and used it in creating the TurVoST. The TSCorpus uses the CQPWeb interface, is freely available, and has been used in other studies prior (e.g., Bilgin 2016). The TSCorpus uses a morphological disambiguator with perception algorithm for Turkish texts for morphological analyses and lemmatization (Sak et al. 2008). The corpus consists of three major Turkish newspapers and a general sampling of Turkish websites. Of the 491 million words, roughly 184 million are from the three newspapers (Radikal, Milliyet, and NTVmsbnc), and 239 million words are from the general sampling of Turkish websites. Sak et al. (2008) explain how these texts were cleaned in detail.

The TSCorpus was preferred over the Turkish National Corpus (TNC) (Aksan et al. 2012), another reliable corpus with 50 M words, for several reasons. First, the TSCorpus is publicly available and therefore users can download the corpus with frequency information included as.txt files or in other formats. Second, the TSCorpus has lemmatization and POS tagging unlike the TNC. Finally, the TNC could not provide information regarding the frequency band of a linguistic item, which would have been a major obstacle in sourcing the items for the TurVoST. One counter argument to using the TSCorpus would be that it lacks a spoken subcorpus. Despite this drawback, the TSCorpus is arguably a viable option because sourcing words from only a written corpus at higher frequency bands would ensure obtaining more low-frequency words. Furthermore, because most Turkish people do not read print materials but read online sources, sourcing vocabulary items from a web and newspaper-based corpus is feasible with the widespread availability of mobile devices and online newspapers.

Developing the Turkish Author Recognition Test (TART)

To construct the TART, a list of popular fiction best-seller author names in Turkey and in Turkish were sourced from Goodreads (Keser 2016). Fiction as a genre was chosen in line with the findings of Wimmer and Ferguson (2022) who demonstrated that fiction-ARTs have the highest explanatory power in vocabulary performance. That is, among ARTs with author names from different genres (fiction, non-fiction, book-counting), the ART with author names of fiction books proved to explain the highest variance in a receptive vocabulary task. In the TART, there were a total of 148 names. Within these 148 names, there were 57 foreign author names. These names were placed on a Google Forms sheet and piloted with 170 participants who were instructed to tick all the authors whose books in Turkish as original or a translated work they recognized as having read. Participants in the piloting were pursuing an

undergraduate (BA) degree or had completed their BA degree at the time of the piloting. Other levels of education were a master's degree (MA) and a doctorate (PhD). After the crowd-sourcing phase, the top 80 scoring author names, which were also known to 50% or more of the participants, were retrieved from the first piloting. Real names were combined with foil names. Foils consisted of Turkish and foreign nonexistent names, each equal to the ratio in the real name list, and they were proportionately gender and nationality matched. Foils were cross-checked to ensure that they did not exist. As with previous ARTs, the maximum score was 80. However, each foil deducted 2 points because of the way real and foil names were counterbalanced (see also Acheson et al. 2008; De la Garza, in preparation). The validation study was conducted in line with the TurVoST to see if the TART would correlate well with the vocabulary size of L1 Turkish speakers. The TART proved to be highly reliable as it demonstrated a Cronbach's alpha of 0.99 and a split-half reliability score (Spearman-Brown corrected) of 0.90.

Because the TART has a binary nature, we calculated a delta prime score for each participant in the TART. Delta prime scores provide insight on participants' ability to discriminate between two categories (i.e., real or foil authors). Following Dimitrov (2016), we calculated correct hits (i.e., selecting real author names), misses (i.e., not selecting real author names), false alarms (i.e., selecting foil author names), and correct rejections (i.e., not selecting foil author names). Then, we utilized the psycho package in R (Makowski 2018) to automatically calculate the delta prime (d') score for each participant. A d' prime score closer to 0 indicates that the participant cannot discriminate between real and foil author names. A d' prime score closer to 3 indicates that the participant is at ceiling in discriminating real and foil author names.

Developing the Turkish Vocabulary Size Test (TurVoST)

The TurVoST is a decontextualized test consisting of 60 items. Each 10 question is from a 1000 word level, sampled from the Turkish Corpus (Sezer and Sezer 2013), such that questions 1–30 test the frequency bands between 3000 and 7000. This is because from a usage-based perspective L1 and L2 speakers alike are expected to know more low-frequency items if they have been exposed to more language, for instance by means of print exposure. The TurVoST follows a multiple-choice question format, and the distractors (excluding the target item) consist of the words from the respective frequency band. The stem consists of words from the first 500 most frequently occurring words. Similarly, the stem of the question is non-defining, and only provides information about the part of speech of the target item. Following the original VST, it does not have an *I don't know* option, and the correct responses are scattered across distractors equally as much as possible. The questions increase in difficulty and the test takes around 10 min to complete. Scoring the test is handled by adding or deducting 1 point to each correct or incorrect response, respectively. It is aimed to be conducted in combination with an ART or batteries of print exposure measures to provide reliable insights on the relationship between literacy and vocabulary knowledge, and other linguistic knowledge (i.e., collocations, grammar).

The carrier sentences indicated the word class to the participants. This was done to ensure that the participants would not experience test-fatigue as this would be used

in combination with other linguistic tasks. This decision also ensured that it could be administered quickly online or in-person, as recruiting volunteers to participate in online studies is difficult. (1) displays an example for an item from the TurVoST.

(1) Tuvalet: Tuvalet-ler temiz

Toilet: Toilet-PL clean
Toilet: toilets are clean

a) Araç-lar-ın git-me-si-ni sağla-yan yuvarlak bir obje (*item: tekerlek*)
Vehicle-PL-GEN go-NOM-POSS able-REL circle a object
"a circular object that helps move vehicles"

b) Hayvan-lar-ın yetiştir-il-diğ-i bir alan (*item: ahır*)
Animals-PL-GEN grow-PASS-REL-POSS a area
"an area in which animals are taken care of"

c) Sür-ül-ebil-en tekerlekli bir araç (*item: motosiklet*)
Drive-PASS-ABLE-REL with.tires a vehicle
"a drivable vehicle with tires"

d) Insan dışkı-sı-nın ve idrar-ı-nın boşaltıl-dığ-ı yer (*item: tuvalet*)
Human feces-ACC-GEN and urine-ACC-GEN empty-PASS-REL-POSS area
"a place where feces and urine can be dumped into"

The original VST contains 140 items, but the TurVoST is an abridged version of it. We removed the first two frequency bands and only sourced items from the odd-numbered frequency bands. Therefore, the TurVoST has 10 vocabulary items each from the 3000, 5000, 7000, 9000, 11,000, and 13,000 frequency bands, selected randomly by the researcher at every 100 words, thus there would be 1 item each from 3000, 3100, 3200, 3300 and so on until 3900. This was repeated for the above-mentioned frequency bands. Only content words were selected, and function words were skipped. Differently from the original VST, we used lemmas as the corpus engine did not provide a frequency list in word families. To the researcher's knowledge, there are currently no Turkish corpora that can provide word families.

To sample the items from the Turkish Corpus, the entire corpus was downloaded as lemmatized in.txt format, thus the frequency order of the items were based on lemmas. Then, ten items from each frequency band as specified above were picked at 1:100 sampling rate (i.e., 1 word in each 100 words). The distractors were created using words from the same frequency band. Differently from the original VST, the TurVoST uses lemmas as word counts for its items. The corpus engine did not allow for word families at the time of the study. The test was first piloted with 15 university lecturers for feedback. There were no issues with the version and thus the study was cleared for a second round of piloting. Following this, 35 L1 Turkish speaking BA students participated in the second round. Item-discrimination scores

lied between 0.51 and 0.78. 54% of the items had an item-discrimination score of 0.51–0.56, 8.64% of the items had a score between 0.56 and 0.62, another 8.64% lied between 0.62 and 0.68, and 2.46% of the questions had an item discrimination score between 0.74 and 0.78. The questions were organized in the frequency order that they were sourced. Cronbach's alpha for the TurVoST was 0.74, suggesting that the test items have reliable construct validity. Split-half reliability (Spearman-Brown corrected) was calculated as follows: we took a random half of each frequency band and had two random halves. The split-half reliability was 0.70, which is well within the acceptable range (Nunnally and Bernstein 1994).

Participants

The study was conducted online on Google Forms and 81 participants (27 females, mean age: 35.41, standard deviation: 14.07) were recruited on a voluntary basis from Turkish university Facebook groups. However, there were students who were enrolled at a university as well as speakers who were not attending university at the time of the study. All speakers were native Turkish speakers and were studying in Ankara at the time of the study. No information on major or dialect were collected as these two variables would not influence the outcome.

Procedure

Participants were first informed about the study and asked for their consent. Then, their background information for age and highest attained degree was collected. Following this, participants were instructed to fill out a self-reported reading questionnaire. A similar questionnaire was used in Dąbrowska (2014) in addition to an author recognition task. The self-reported reading questionnaire has a total of 5 questions. While four questions ask for the same information, two of them are formulated to see how much they read in work or school contexts, and the other two ask how much they read on their own initiative. Participants are asked to rate how often they read emails, messages, newspapers, books, scientific articles, social media, comic books among others (see appendix). The final question asks participants to rate statements about reading. All answers were transformed into numerical codes, 0 through 6, with higher numbers indicating more frequency of reading and 0 indicating never. Scoring on the final question was reversed for negative answers, with 0 through −6. The maximum score a person could obtain on the print exposure questionnaire was 180. In the end, the scores were summed to create the 'print exposure' score. The split-half reliability (Spear-Browman corrected) was found to be 0.89.

Following this, participants were presented with the TART, in which all the author names were randomized. They were instructed to mark the names they were familiar with, including foreign authors whose works they may have read as translations. Then, questions in the TurVoST was presented in a non-randomized order.

Data analysis

The effect of the four predictors (highest attained degree, age, author recognition task, self-reported questionnaire) on the scores of TurVoST was calculated and measured by using standard multiple regression modeling in R (R Core Team 2021), the codes used are available in appendix. Descriptives were visualized using boxplots prior to regressions. The "degree" variable, representing the educational attainment of participants, was encoded as a categorical variable to facilitate statistical analysis. Initially stored as numeric values ranging from 1 to 4 denoting different levels of education (1 for high school, 2 for BA, 3 for MA, and 4 for PhD), the variable underwent a transformation to ensure appropriate labeling. This transformation was achieved using the factor() function in the R programming language. This ensured its usability in the regression model. The initial model contained all four predictors (age, degree, TART and print exposure as main effects) and the subsequent interactions. Nonsignificant predictors were removed one after the other after running a log likelihood test, starting with the highest interaction term with the largest p-value, as suggested by Crawley (2010). If the predictor did not improve the fit of the model, it was removed. This process was done for all the predictors until all predictors were significant. All predictors except for degree were centered using the scale() function to facilitate interpretation of results. To interpret the model coefficients more effectively, the relative importance of each predictor was measured using the lmg metric, computed by the relaimpo package in R (see Grömping 2007). This metric is calculated by averaging the sum-of-squares obtained from all possible orderings of the predictors in the model, and is thought to be analogous to a squared semi-partial correlation. Larson-Hall (2010) argues that it quantifies the variance explained by each predictor in the model. Model assumptions were checked using the performance package in R (Lüdecke et al. 2021) Heteroscedasticity was detected in the model. We used the Eicker-Huber-White method to adjust the values for standard errors, *t*-values and *p*-values (White 1980).

Results

Relationship between reading and vocabulary size

Table 2 outlines the descriptive statistics for the data of 81 participants and Fig. 1 visualizes age, TART, and self-reported reading scores. Figure 2 visualizes individual scores attained on the TurVoST. As is clear, the age of the participants varied between 17 and 69, with a mean of 35 years of age. Because of the way in which information on degree was collected, it can be concluded that the majority of the participants had completed a BA degree, although there were also participants with a high school, master's or a PhD degree (1 indicates a high school degree, 2 undergraduate "BA", 3 masters "MA", and 4 doctorate "PhD").

Results on the vocabulary size test show that on average participants answered 55.63 out of 60 questions correctly. Self-reported reading (i.e., results from the print exposure questionnaire) show that the participants had a mean reading score of 76 out

Table 2 Descriptive statistics

	TART	TART d' Prime	Vocabulary	Age	Degree	Self-reported reading
Mean	62.14	−1.02	55.63	35.42	2.18	76.04
Median	68	−1.11	56	32	2	77
Standard Deviation	16.07	1.02	3.51	14.07	0.93	19.09
Minimum	13	−3.11	43	17	1	24
Maximum	79	2.28	60	69	4	115
25th Percentile	48	−1.56	54	24	1	66
75th Percentile	75	−0.55	58	50	3	89

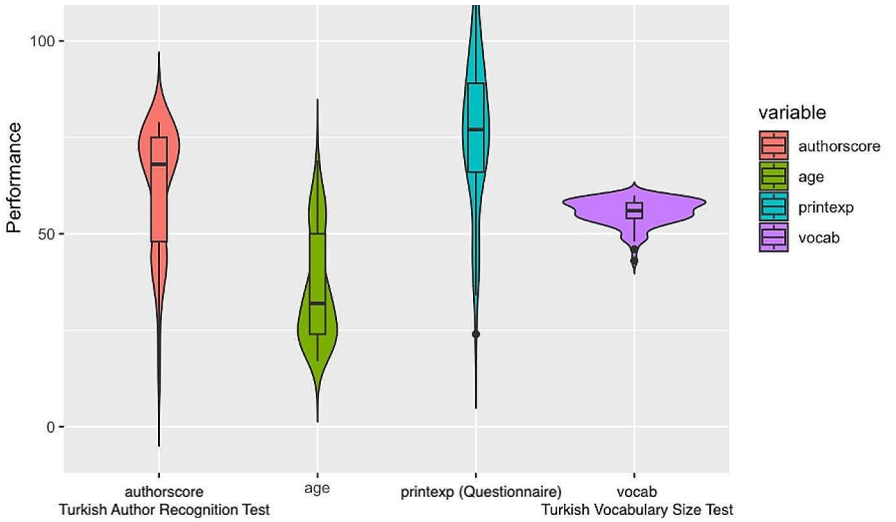

Fig. 1 Performance on the TurVoST (percent correct). *Abbreviations authorscore* number of correct answers in the TART, *printexp* number of quantified responses in the self-reported reading questionnaire

of 180. The scores for TART show that the participants on average had read or could recognize 62 real authors in Turkish out of 80 real authors. TART d' scores indicate that on average the participants were at chance in discriminating against real and foil authors. This approach, however, needs to be taken with precaution as the participants were only asked to indicate the authors whose work they had read previously. They were not instructed to reject (or check) foil authors, which would have been the case in a lexical-decision like design. Therefore, the design in which the author names are presented may interfere with d' prime scores.

Table 3 outlines pairwise correlations. Pearson correlations point to a statistically significant ($p<0.0001$) and strong correlation between the accuracy on the TurVoST and the TART (r=0.47). Using d' prime to ensure the binary nature of the TART would not interfere, the correlation was significant ($r=0.39$, $p=0.0003$), albeit lower than the correlation between the TART and the TurVoST. Age also correlated strongly with vocabulary scores at 0.50 ($p=0.0000$). Education or the highest attained degree also correlated with vocabulary knowledge at 0.41 ($p=0.0002$). Interestingly, the questionnaire for self-reported reading did not correlate with any variables in the

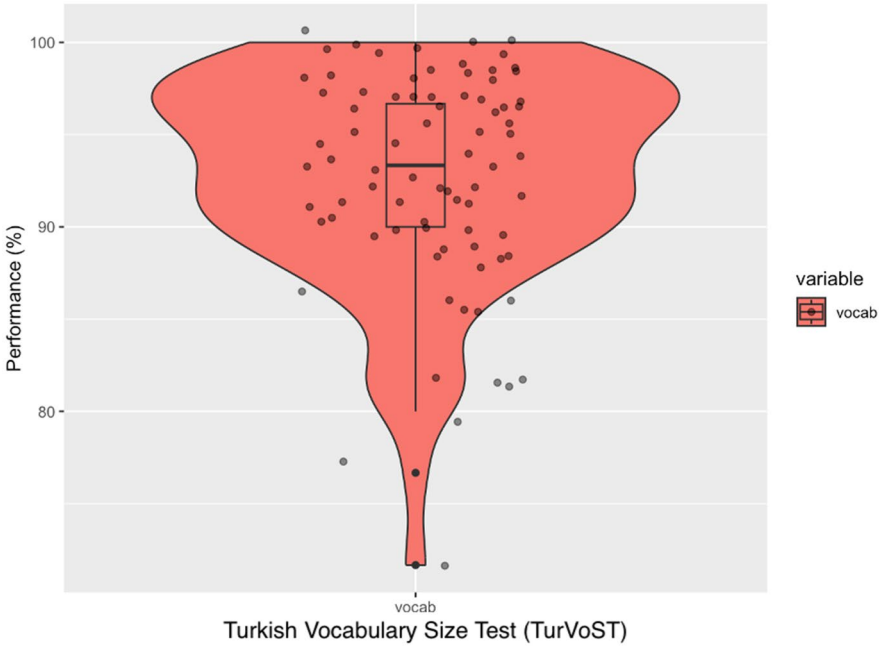

Fig. 2 Performance in the Turkish Vocabulary Size Test (TurVoST) with individual observations

study. Therefore, these results provide further suggestive evidence that people provide socially desirable answers on reading exposure questionnaires (e.g., Acheson et al. 2008). If this had not been the case, then the questionnaire scores would have correlated with vocabulary and other variables. Figures 3 and 4 visualize the relationship between vocabulary and the TART, and age and vocabulary, respectively.

The results for the final best fitting model fit are shown in Table 4. The model explains a statistically significant and substantial proportion of variance ($R^2 = 0.49$, $F(6, 74) = 12.02$, $p < 0.001$, adj. $R^2 = 0.45$). The model's intercept, corresponding to TART = 0, degree = High School and age = 0, is at −0.27, $t(74) = -1.51$, $p = 0.250$.

Table 3 Pairwise correlations between variables

	TART	Age	Degree	Self-reported reading	Vocabulary	TART d' Prime
TART	1.00					
Age	0.19	1.00				
Degree	0.08	0.37*	1.00			
Self-reported Reading	−0.05	−0.11	0.15	1.00		
Vocabulary	0.47***	0.50***	0.41**	0.11	1.00	
TART d' Prime	0.70****	0.12	0.01	−0.11	0.39*	1.00

Asterisks indicate different levels of statistical significance: $p < 0.0001$ '****'; $p < 0.001$ '***', $p < 0.01$ '**', $p < 0.05$ '*'

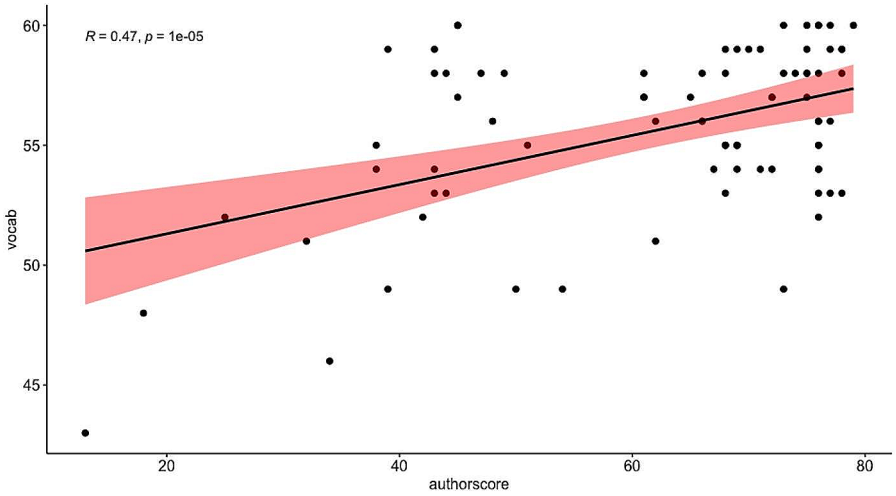

Fig. 3 The relationship between the Turkish Author Recognition Test (TART) and vocabulary scores in the Turkish Vocabulary Size Test (TurVoST)

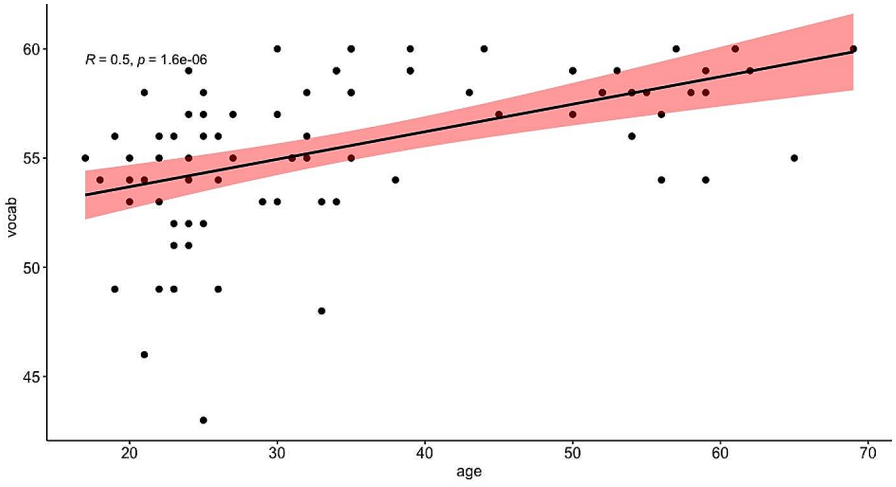

Fig. 4 The relationship between age and vocabulary scores in the Turkish Vocabulary Size Test (TurVoST)

Due to heteroscedasticity, we used the Eicker-Huber method to correct the *t* and *p* values, as explained in the methods section. As is seen in the results in Table 4, age and TART account for roughly the same amount of variance at approximately 18%. Degree (all BA, MA and PhD) accounts for roughly another 9%, and there is a significant interaction between reading and age, accounting for variance at 3.75%. As can be seen from the degrees, higher levels of education result in more vocabulary knowledge. As TART increases, vocabulary size tends to increase, as indicated by the positive coefficient for TART. However, this positive effect of TART on vocabulary

Table 4 Final best fitting model

Variable	Estimate	Standard error	t value	Pr(>|t|)	lmg
Intercept	–0.206	0.178	–1.15	0.250	
TART	0.319	0.089	3.57	0.000	0.1773
degreeBA	0.160	0.243	0.659	0.511	0.0982
degreeMA	0.464	0.241	1.92	0.058	
degreePhD	0.706	0.345	2.04	0.044	
age	0.406	0.099	4.07	0.000	0.1807
TART*age	–0.235	0.096	–2.45	0.016	0.0373
Adjusted Model R2: 0.45					

size decreases as age increases, as indicated by the negative coefficient for the interaction term between TART and age.

Figure 5 visualizes this interaction: if speakers do not have a great deal of print exposure and are young, that seems to be particularly detrimental for vocabulary knowledge (i.e., found in the lower right quadrant marked with bright yellow). Similarly, if a speaker is 40 years old or above, but does not have much experience with written language, then age can compensate for a lack of reading exposure (i.e., the brown dots placed in the 40–60 range for TART). Finally, speakers regardless of age who read a lot seem to demonstrate close to ceiling performance on vocabulary (i.e., bright yellow dots found in the upper right quadrant).

Results for the TurVoST

Figure 6 visualizes mean accuracy scores in each frequency band. As is clear, with growing frequency (i.e., with lower frequency words), mean accuracy decreases with growing frequency bands. Figure 7 visualizes mean accuracy scores per frequency band:

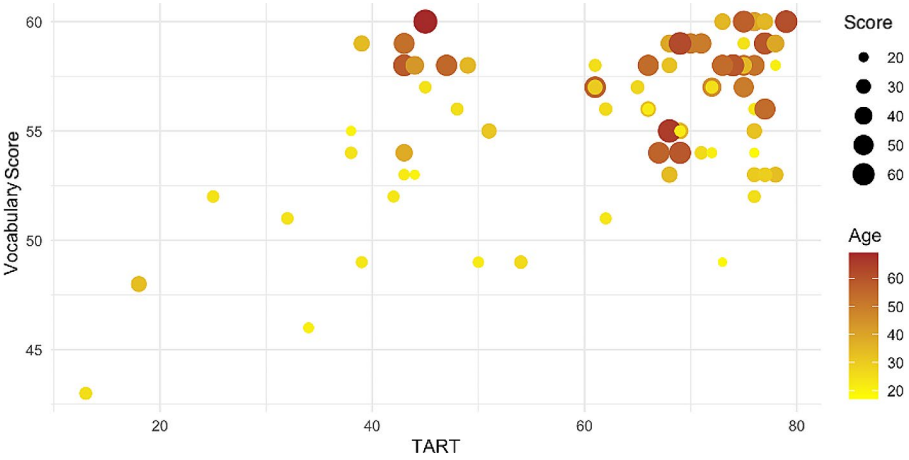

Fig. 5 The relationship between vocabulary scores, performance on the Turkish Author Recognition Test (TART), and age

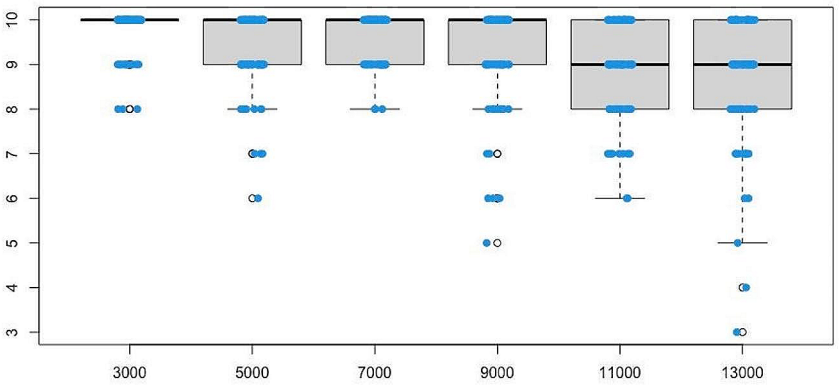

Fig. 6 Accuracy scores in the Turkish Vocabulary Size Test (TurVoST) with individual points

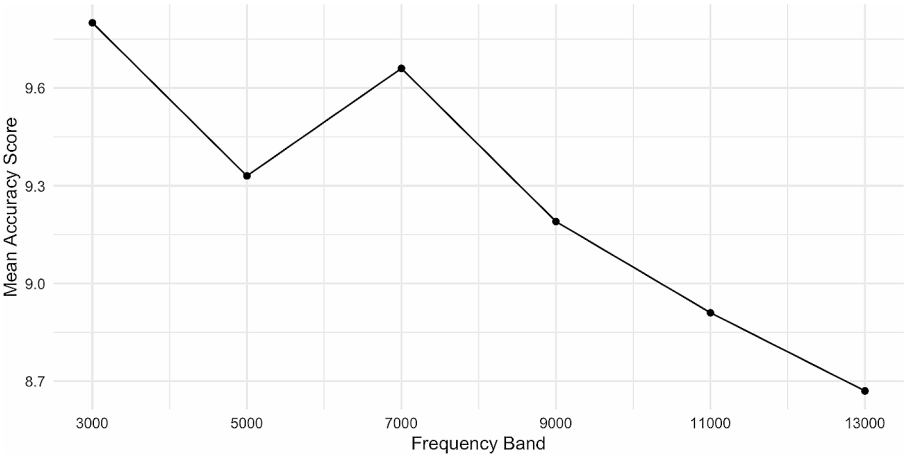

Fig. 7 Mean accuracy across frequency bands (maximum score 10 per frequency level) in the Turkish Vocabulary Size Test (TurVoST)

- at 3000 speakers were 98% correct,
- at 5000 this was 93.30%,
- at 7000 this increased to 96.60%,
- at 9000 this dropped to 91.89%,
- at 11,000 this dropped further down to 89.1%,
- and at 13,000 it was 86.7%.

Unsurprisingly, the 13,000 frequency band demonstrates the most individual differences. On average, participants had an average accuracy of 55.63. In comparison, Dąbrowska's participants (2018) accurately responded to 41.4 vocabulary items out of 60. This is quite lower than the current study. One potential explanation for this might be that Dąbrowska's study (2018) included people from both high and low aca-

demic backgrounds, and this heterogeneity may have resulted in such a difference. Whereas in the present study, the majority of the participants are of high academic attainment backgrounds (i.e., they have completed their BA degrees).

Discussion and implications

Although the results in this paper do not necessarily reinvent the wheel, they are important for Turkish linguistics as well as for usage-based approaches to language learning for two reasons. First, this is the first study to examine literacy-related individual differences in vocabulary knowledge among L1 Turkish speakers with two novel research instruments. Second, from a theoretical perspective, our findings provide further converging evidence in favor of usage-based approaches from an understudied language within the framework. The findings suggest that more exposure to language increases vocabulary knowledge.

Our findings are consistent with previous research, demonstrating a relationship between reading and vocabulary knowledge as outlined earlier in the paper. This seeming facilitation of reading is because reading modulates the exposure to language, giving speakers more opportunities to encounter a more enriched language experience (e.g., Huettig and Pickering 2019). This resonates with Dąbrowska's (2009) maxim that most vocabulary learning occurs incidentally in adulthood through reading is reinforced with the findings of this study.

The lack of correlation between the TART and the results from the self-reported literacy questionnaire, as well as the TurVoST and the questionnaire is interesting. This is likely because responses to self-reported literacy questionnaires are inflated due to social desirability, as most of the questions in the questionnaire prompt answers to 'how many hours do you read X?' (e.g., Acheson et al. 2008). As such, the responses given in the questionnaire may not potentially reflect the real-life reading results of these participants, and therefore the results from the questionnaire may not have correlated with the scores of the TART or the TurVoST.

In the TurVoST, accuracy rates steadily decreased from 98% in the first three-thousand-word frequency band to 89% in the thirteen thousand frequency band. This finding is in line with usage-based linguistics and its assumption that lower frequency linguistic items are more difficult to access reliably as a result of a lack of exposure (e.g., Bybee 2010). Thus, it is not surprising to see that reading exposure or maturation were independent predictors of determining minute variations in accuracy scores in the TurVoST. In terms of individual differences in vocabulary knowledge, most participants performed at ceiling. However, the performance of those that did not perform at ceiling, especially in the 11,000 and 13,000 frequency bands, were well predicted by print exposure and maturation.

Exposure through age is another vital aspect in addition to reading. Different subgenres of spoken language can contain lexicogrammatically enriched language (Biber 2009). As such, as a person matures, they may have more opportunities to be exposed to different varieties of texts and spoken modalities. Therefore, age increases the probability of learning more vocabulary incidentally.

A skeptic might note that most of the participants in this study performed at ceiling on the vocabulary task. While this might raise concerns about how well the Tur-

VoST measures receptive vocabulary size in Turkish, it is a valid task when used in combination with other linguistic measures. We propose two explanations. First, there was a strong and significant correlation between reading and vocabulary knowledge—even under ceiling effects, suggesting that exposure can account for a large proportion of variances in a population that showed relatively little variation in the TurVoST. This is not surprising. Comparing these results against Dąbrowska (2018), the ART scores in her study accounted for about 25% of the variance in vocabulary scores, whereas in the current study it is about 17%. This difference is likely the result of a more diverse population from different educational backgrounds in Dąbrowska's study. Such that the mean score on the vocabulary test in Dąbrowska (2018) was 41 whereas in the current one it was 55. Therefore, more variation to be explained will also result in a higher percentage of variation explained. Secondly, accuracy scores across growing frequency bands in the TurVoST provide further converging evidence to the idea that language learning, in this case vocabulary, is modulated by frequency, and provides more experimental evidence in favor of usage-based approaches to language learning. And as such, the variance accounted for by the TART, which is predominantly found in the upper frequency levels of the TurVoST, explains that those who read more are exposed to more Turkish, and as such incidentally learned more vocabulary items than those that read less.

The fact that approximately 18% of the variance in vocabulary knowledge of those that did not perform at ceiling can be captured by the TART aligns with our expectations. Moreover, this trend illuminates a fundamental aspect of lexical development: the notion that extensive reading, over time, can lead individuals of varying ages and educational backgrounds to attain comparable levels of lexical proficiency. Similarly, it suggests that elderly individuals who do not read as much may attain a similar performance of vocabulary (either as a result of mere exposure or education).

This observation is underpinned by the pervasive nature of written-language-biased vocabulary items, which, although predominantly encountered in written form, permeate spoken language to a considerable extent. For instance, language written to be spoken often exhibits greater lexical richness compared to those employed in spontaneous speech, such as news broadcasts (Biber 2009). Our study serves as an illustration of this. Our results also show that attained degree (i.e., education) can influence vocabulary knowledge. This is not surprising: education provides opportunities for reading and enables speakers to be engaged in highly literate circles whereby linguistic encounters may have more low-frequency vocabulary items. We provide evidence in support of usage-based approaches to language acquisition, wherein individuals who engage in extensive reading are exposed to a broader array of lexical items, including those categorized as "low-frequency" or predominantly encountered in written contexts. They also provide further evidence in favor of usage-based approaches that language learning is a journey that is dynamic, emergent, and a result of exposure. At the end of the day, the TART and the TurVoST measure what they are designed to measure efficiently: the effects of reading on vocabulary knowledge.

We compiled Tables 5 to provide a more comprehensive overview of the relationship between reading and its effects on vocabulary knowledge. Table 5 compares the correlations between vocabulary scores and scores on various ARTs. As is seen in the table, previously reported correlations lie anywhere between 0.35 and 0.62, and the

Table 5 Correlations between scores on various ARTs and vocabulary knowledge

Study	Number of participants	Test used	VST reliability (Cronbach)	Correlation	Art used	ART reliability (Cronbach)
Current study	81	Turkish Vocabulary Size Test (TurVoST)	0.74	0.47	Turkish ART	0.99
Brysbaert et al. (2020)	195	Vocabulary (Dutch Vocabulary Test)	0.87	0.42	Dutch ART (Brysbaert et al. 2020)	0.95
Dąbrowska (2018)	90	Vocabulary (VST)	0.91	0.60	English ART (Acheson et al. 2008)	–
De la Garza (in preparation)	100	Vocabulary (Lex-TaleESP)	0.96	0.37	Spanish ART (De la Garza, in preparation)	0.80
James et al. (2018)	123	Vocabulary (Extended Range Vocabulary Test)	–	0.45	English ART (Acheson et al. 2008)	–
Lee et al. (2019)	105	Vocabulary (self developed)	–	0.35	Korean ART (Lee et al. 2019)	0.99
Payne et al. (2012)	139	Vocabulary (Educational Testing Service Kit of Factor Referenced Cognitive Tests)	–	0.62	English ART (Stanovich and West 1989)	0.84

– indicates information not available

TART and the TurVoST fit well into it with a correlation of 0.47. Similarly, when the correlation between the TART d' scores and vocabulary knowledge is analyzed, there appears to be a significant correlation ($r = 0.39$, $p < 0.0001$). This shows that there is not a meaningful difference between vanilla TART scores or TART d' scores in terms of predictive power. Importantly, however, this lack of difference between vanilla and d' scores of ARTs seem to manifest itself in checkbox-style designs: online or pen & paper (e.g., Wright et al. in preparation). Similarly, the reliability scores for both novel research instruments rely well within the range reported for other instruments in the field or within the acceptable range, i.e., TurVoST. The correlations in light of previous research studies, combined with the reliability scores of the TART and the TurVoST, leads us to conclude that the TART can assess print exposure and the TurVoST can measure vocabulary knowledge reliably and can be used as a research tool with high explanatory power for literacy-related individual differences in linguistic studies in Turkish.

As can be seen from our discussions and Table 5, our results are consistent with previous studies that established a moderate to strong correlation between vocabulary knowledge and reading measured by the TART and vocabulary scores among highly educated populations (e.g., Burt and Fury 2000; Grant et al. 2007; Stanovich and Cunningham 1992; Stanovich et al. 1995; West and Stanovich 1991). Vocabulary knowledge, measured by the TurVoST, also provides interesting findings. The participants in this study on average demonstrate having the receptive knowledge of roughly 55/60 words. Comparing the results of Dąbrowska (2018), it appears that on average an L1 English sample from the UK on average can correctly respond to 41/60 questions in the English VST. Considering that Dąbrowska's participants were from a diverse background of academic attainment, it is safe to claim that the Turkish participants in this study outperform them. This reiterates the importance of education. Higher academic attainment usually implies reading more print materials, or being in social circles that read often, both of which lead to lexicogrammatical enrichment of the ambient language (cf. Dąbrowska 2018), which is reflected in the regression analyses. As such, speakers that are engaged in either of these circumstances appear to have a more diverse vocabulary knowledge. This is also reflected in our findings, where the highest attained degree accounts for another 10% of variance in the results of the TurVoST.

Limitations and further research

Currently, the two research instruments have only been tested with a small population. This study only reported the development of the two research instruments. Future studies should validate these instruments, especially the TurVoST using established methods of analyzing validation. To validate both the reliability of each instrument and the correlation between the two, further studies need to be conducted on more diverse Turkish speaking adult populations (i.e., illiterates, low-literates). Because the TurVoST only has 60 items, it is quite difficult to establish vocabulary size, as is the case in the original VST as it has 140 items. Future work can also include more difficult items from higher frequency bands to increase the sensitivity of this task. This was because the TurVoST has been mainly designed to be used for

linguistic research in combination with other language tasks rather than establishing vocabulary size of L1 Turkish speakers.

Future research should incorporate a broader range of demographic data to better understand these influences. We recommend gathering detailed information on participants' dialect, ethnicity, nationality, and other relevant factors to explore how these variables may affect linguistic abilities and literacy outcomes, relevant for vocabulary knowledge. Dialectal differences, ethnicity, nationality, and race may influence vocabulary acquisition by determining access to education and reading.

Future studies should consider this in mind. In addition, the TurVoST should be validated with more L1 Turkish speakers to observe if ceiling effects persist, in which case the TurVoST can be improved upon by adding more items from further frequency levels. Furthermore, both instruments were designed with Turkey Turkish in mind. However, the TART and the TurVoST also need to be validated with heritage speakers of Turkish, and other dialects of Turkish (i.e., Cypriot Turkish). Finally, both tools were developed with L1 research in mind. To ensure that the TART and the TurVoST work in L2 learners of Turkish, they need to be validated with L2 speakers of Turkish from various proficiency levels.

Conclusion

This study reported the development of the first Turkish Author Recognition Test (TART) and the Turkish Vocabulary Size Test (TurVoST). Many previous studies point to a positive correlation between reading, as measured by author recognition tasks, and vocabulary knowledge (e.g., Cunningham and Stanovich 1998; Dąbrowska 2018). However, this had not yet been operationalized in a Turkish L1 speaking adult population. Our findings show acceptable reliability scores for both tests and point to a statistically significant correlation between the TART and the TurVoST ($r = 0.47$). The study also uncovered the effects of maturation on vocabulary knowledge. Such that, in a regression model, the TART accounts for almost 18% of the variation in the vocabulary scores, followed by age at about 17%, and highest attained degree at 10%. The study provides the first account of the relationship between reading and vocabulary knowledge among adult L1 Turkish speakers and demonstrates that print exposure modulates vocabulary size alongside other factors such as education and maturation, another proxy for language exposure, to a very large extent.

Acknowledgements I would like to thank everyone who participated in the study as well as the two reviewers for their diligent feedback. I would like to thank Fatıma Uslu for her help during data collection and preparing the materials. All remaining errors are mine.

Author contributions The author was responsible for everything.

Funding This study was not funded.
Open Access funding enabled and organized by Projekt DEAL.

Data availability It is available in the link in the abstract.

Declarations

Ethical approval The research was performed in accord with the ethical guidelines prescribed by the Helsinki Declaration regarding human participants in research.

Informed consent Informed consent was obtained from all study participants.

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Acheson DJ, Wells JB, MacDonald MC (2008) New and updated tests of print exposure and reading abilities in college students. Behav Res Methods 40(1):278–289. https://doi.org/10.3758/BRM.40.1.278

Aksan Y, Aksan M, Koltuksuz A, Sezer T, Mersinli Ü, Demirhan UU, Yılmazer H, Atasoy G, Öz S, Yıldız İ, Kurtoğlu Ö (2012) Construction of the Turkish National Corpus (TNC). In: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12), pp 3223–3227. http://www.lrec-conf.org/proceedings/lrec2012/pdf/991_Paper.pdf

Beglar D (2010) A rasch-based validation of the vocabulary size test. Lang Test 27(1):101–118

Berument SK, Güven AG (2013) Türkçe İfade Edici ve Alıcı Dil (TİFALDİ) Testi: I. alıcı Dil Kelime alt testi standardizasyon ve güvenilirlik geçerlik çalışması. Turk Psikiyatri Derg 24(3):192–201

Biber D (2009) Dimensions of register variation. Cambridge University Press

Bilgin O (2016) Frequency effects in the processing of morphologically complex Turkish words. Master's Thesis, Boğaziçi University

Burt JS, Fury MB (2000) Spelling in adults: the role of reading skills and experience. Read Writ 13(1/2):1–30. https://doi.org/10.1023/A:1008071802996

Brysbaert M, Sui L, Dirix N, Hintz F (2020) Dutch author recognition test. J Cognit 3(1). https://doi.org/10.5334/joc.95

Bybee J (2010) Language, usage and cognition. Cambridge University Press. https://doi.org/10.1017/CBO9780511750526

Chen S-Y, Fang S-P (2015) Developing a Chinese version of an author recognition test for college students in Taiwan. J Res Read 38(4):344–360. https://doi.org/10.1111/1467-9817.12018

Chomsky N (2014) Aspects of the Theory of Syntax (No. 11). MIT press.

Crawley M (2010) R a language and environment for statistical computing: reference index. R Foundation for Statistical Computing

Cunningham AE, Stanovich KE (1998) What reading does for the mind. Am Educ 22:8–17

Dąbrowska E (2009) Words as constructions. In: Evans V, Pourcel S (eds) Human cognitive processing, vol 24. John Benjamins Publishing Company, pp 201–223. https://doi.org/10.1075/hcp.24.16dab

Dąbrowska E (2014) Words that go together: measuring individual differences in native speakers' knowledge of collocations. Ment Lex 9(3):401–418. https://doi.org/10.1075/ml.9.3.02dab

Dąbrowska E (2015) Individual differences in grammatical knowledge. In: Dąbrowska E, Divjak D (eds) Handbook of cognitive linguistics. De Gruyter Mouton, pp 650–668. https://doi.org/10.1515/9783110292022-033

Dąbrowska E (2018) Experience, aptitude and individual differences in native language ultimate attainment. Cognition 178:222–235. https://doi.org/10.1016/j.cognition.2018.05.018
Dąbrowska E (2020) Language as a phenomenon of the third kind. Cogn Linguist 31(2):213–229. https://doi.org/10.1515/cog-2019-0029
De la Garza V (in preparation) Spanish Author Recognition Task
Dimitrov DM (2016) An approach to scoring and equating tests with binary items: piloting with large-scale assessments. Educ Psychol Meas 76(6):954–975. https://doi.org/10.1177/0013164416631100
Divjak D (2019) Frequency in language: memory, attention and learning. Cambridge University Press
Eisenbeiß S (2009) Generative approaches to language learning. Linguistics 47(2). https://doi.org/10.1515/LING.2009.011
Goldberg AE (2006) Constructions at work: the nature of generalization in language. Oxford University Press
Goldberg AE (2019) Explain me this: creativity, competition, and the partial productivity of constructions. Princeton University Press
Grant A, Wilson AM, Gottardo A (2007) The role of print exposure in ReadingSkills of Postsecondary Students with andWithout reading disabilities. Exceptionality Educ Int 17(2). https://doi.org/10.5206/eei.v17i2.7603
Grömping U (2007) Estimators of relative importance in linear regression based on variance decomposition. Am Stat 61(2):139–147. https://doi.org/10.1198/000313007X188252
Gyllstad H, McLean S, Stewart J (2021) Using confidence intervals to determine adequate item sample sizes for vocabulary tests: An essential but overlooked practice. Language Test (4):558–579. https://doi.org/10.1177/0265532220979562
Hayes DP, Ahrens MG (1988) Vocabulary simplification for children: a special case of 'motherese'? J Child Lang 15(2):395–410
Huettig F, Pickering MJ (2019) Literacy advantages beyond reading: prediction of spoken language. Trends Cogn Sci 23(6):464–475. https://doi.org/10.1016/j.tics.2019.03.008
James AN, Fraundorf SH, Lee E-K, Watson DG (2018) Individual differences in syntactic processing: is there evidence for reader-text interactions? J Mem Lang 102:155–181. https://doi.org/10.1016/j.jml.2018.05.006
Katz J et al (1974) A Turkish peabody picture vocabulary test. Hacettepe Bull Soc Sci Humanit
Keser H (2016) Popüler Türk Yazarları ve Kitaplar. https://public.tableau.com/app/profile/hasan.keser/viz/Book1_16818/Story1
Kidd E, Donnelly S, Christiansen MH (2018) Individual differences in language acquisition and processing. Trends Cogn Sci 22(2):154–169. https://doi.org/10.1016/j.tics.2017.11.006
Larson-Hall J (2010) A guide to doing statistics in second language research using SPSS and R, Second edn. Routledge
Lee H, Seong E, Choi W, Lowder MW (2019) Development and assessment of the Korean author recognition test. Q J Exp Psychol 72(7):1837–1846. https://doi.org/10.1177/1747021818814461
Lüdecke D, Ben-Shachar MS, Patil I, Waggoner P, Makowski D (2021) Performance: An R package for assessment, comparison and testing of statistical models. J Open Source Softw 6(60). https://doi.org/10.21105/joss.03139
Makowski D (2018) The psycho package: an efficient and publishing-oriented workflow for psychological science. J Open Source Softw 3(22):470
Mar RA, Rain M (2015) Narrative fiction and expository nonfiction differentially predict verbal ability. Sci Stud Read 19(6):419–433. https://doi.org/10.1080/10888438.2015.1069296
McLean S, Hogg N, Kramer B (2014) Estimations of Japanese university learners' English vocabulary sizes using the vocabulary size test. Vocab Learn Instr 3(2):47–55
Nation I, Beglar D (2007) A vocabulary size test. Lang Teach 21(7):9–13
Nunnally JC, Bernstein IH (1994) Psychometric theory, 3rd edn. McGraw-Hill
Payne BR, Gao X, Noh SR, Anderson CJ, Stine-Morrow EAL (2012) The effects of print exposure on sentence processing and memory in older adults: evidence for efficiency and reserve. Aging Neuropsychol Cogn 19(1–2):122–149. https://doi.org/10.1080/13825585.2011.628376
R Core Team (2021) R: a language and environment for statistical computing. R Foundation for Statistical Computing. https://www.r-project.org/
Sak, H, Güngör T, Saraçlar M (2008) TurkishLanguage resources: Morphological parser, morphological disambiguator and web corpus. Proceedings of the 6th international conference onAdvances in Natural Language Processing, Gothenburg, pp. 417–427.

Schmitt N, Schmitt D (2014) A reassessment of frequency and vocabulary size in L2 vocabulary teaching. Language Teach 47(4):484–503. https://doi.org/10.1017/S0261444812000018

Sezer B, Sezer T (2013) TS corpus: Herkes için Türkçe derlem. In: Proceedings of the 27th National Linguistics Conference, pp 217–225

Stanovich KE, Cunningham AE (1992) Studying the consequences of literacy within a literate society: the cognitive correlates of print exposure. Mem Cognit 20(1):51–68. https://doi.org/10.3758/BF03208254

Stanovich KE, West RF (1989) Exposure to print and orthographic processing. Read Res Q 24(4):402. https://doi.org/10.2307/747605

Stanovich KE, West RF, Harrison MR (1995) Knowledge growth and maintenance across the life span: the role of print exposure. Dev Psychol 31(5):811–826. https://doi.org/10.1037/0012-1649.31.5.811

Stewart J, Stoeckel T, McLean S, Nation P, Pinchbeck GG (2021). What the research shows about written receptive vocabulary testing: A reply to Webb. Stud Second Language Acquis 43(2):462–471. https://doi.org/10.1017/S0272263121000437

Stoeckel T, McLean S, Nation P (2021) Limitations of size and levels tests of written receptive vocabulary knowledge. Stud Second Language Acquis 43(1):181–203. https://doi.org/10.1017/S027226312000025X

West RF, Stanovich KE (1991) The incidental acquisition of information from reading. Psychol Sci 2(5):325–330. https://doi.org/10.1111/j.1467-9280.1991.tb00160.x

Wimmer L, Ferguson HJ (2022) Testing the validity of a self-report scale, author recognition test, and book counting as measures of lifetime exposure to print fiction. Behav Res Methods 55(1):103–134. https://doi.org/10.3758/s13428-021-01784-2

White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica J Econom Soc 817–838. https://doi.org/10.2307/1912934