frontiers | Frontiers in Language Sciences

TYPE Original Research
PUBLISHED 24 September 2025
DOI 10.3389/flang.2025.1633610

OPEN ACCESS

EDITED BY
José Antonio Hinojosa,
Complutense University of Madrid, Spain

REVIEWED BY
Laura Muscalu,
Hobart and William Smith Colleges, United States
Qichao Song,
Beijing Foreign Studies University, China

*CORRESPONDENCE
Michael S. Vitevitch
mvitevit@ku.edu

RECEIVED 10 June 2025
ACCEPTED 28 August 2025
PUBLISHED 24 September 2025

CITATION
Aljasser FM, Wang Y and Vitevitch MS (2025) Word-likeness ratings of English-like non-words by Arabic and Chinese learners of English: implications for assessing proficiency. *Front. Lang. Sci.* 4:1633610. doi: 10.3389/flang.2025.1633610

Word-likeness ratings of English-like non-words by Arabic and Chinese learners of English: implications for assessing proficiency

Faisal M. Aljasser [1], Yusen Wang [2] and Michael S. Vitevitch [3]*

[1]Department of English Language and Literature, College of Languages and Humanities, Qassim University, Buraydah, Saudi Arabia, [2]Kunming University of Science and Technology, Kunming, China, [3]University of Kansas, Lawrence, KS, United States

The present study investigated English learners' subjective word-likeness ratings of English-like non-words. For both native Arabic and Mandarin speakers learning English, the subjective word-likeness ratings correlated with objective phonotactic probability measures of those non-words. Although there were no differences between the subjective word-likeness ratings of Beginner and Advanced English learners, the word-likeness ratings of Beginner and Advanced learners did differ from the ratings provided (in a previous study) by native English speakers, suggesting that subjective ratings of phonotactic knowledge may be useful in measuring the extent to which phonological knowledge approaches native-like levels. The results of the present study underscore the important role of phonotactic knowledge in language proficiency. Implications of these findings to second language learning are discussed, with an emphasis on how the non-word rating task might be used as a novel and efficient method to assess one aspect of language proficiency.

KEYWORDS

phonotactic probability, word-likeness ratings, proficiency, assessment, learner intuition

Introduction

Knowledge of the sequences of phonemes that constitute a word in a language is referred to as *phonotactic information* (Vitevitch and Aljasser, 2021). Sensitivity to phonotactic information occurs early in life as demonstrated in studies of word learning where Jusczyk et al. (1993, 1994) first showed that English and Dutch learning infants at 9 m.o. exhibit a preference for novel words that aligned with the phonotactic patterns of their native language. Early sensitivity to the phonotactic patterns of one's native language has been subsequently observed in a wide range of languages including Turkish, French, Catalan, Hungarian, and Japanese (see Sundara et al., 2022, for a review).

Familiarity with the phonotactic patterns of one's native language early in life may influence the later acquisition of words in infants, older children, and adults (e.g., Graf Estes et al., 2011; Storkel and Rogers, 2000; Storkel et al., 2006). For instance, Gonzalez-Gomez et al. (2013) found that 14-month-old French-learning infants were able to learn words with high phonotactic probability in French, whereas 16-month-olds successfully learned both high and low phonotactic probability words. Even in more mature individuals

with larger vocabularies, such as typically developing 3–5-year-old children (Hoover et al., 2010), children who are late talkers (MacRoy-Higgins and Dalton, 2015), and adults (Storkel et al., 2006), phonotactic probability still plays a role in triggering the word learning process. That is, a word with low rather than high phonotactic probability "stands out" as being an item that is absent from the lexicon, which leads to the allocation of cognitive resources and the initiation of processes to learn that novel word.

Knowledge of the phonotactic patterns of a language not only helps in learning new words but also plays a role in segmenting known words from fluent speech. This sensitivity to phonotactic probability for lexical segmentation develops in infancy (e.g., Saffran et al., 1996; Mattys et al., 1999; Mattys and Jusczyk, 2001), and continues to be used in adulthood while listening to L1 (McQueen, 1998; Mersad and Nazzi, 2011) and L2 (Weber and Cutler, 2006; Al-jasser, 2008), where quite subtle differences in phonotactic probability suffice to signal word boundaries (Dal Ben et al., 2021).

Finally, phonotactic probability has also been shown to influence the production and recognition of speech. Studies of speech production have demonstrated influences of phonotactic probability in Dutch-learning (Zamuner, 2009) and English-learning children (Storkel and Maekawa, 2005), and in English speaking adults (Goldrick and Larson, 2008). For evidence of the influence of phonotactic probability on speech recognition see studies by Kelley and Tucker (2022) and Vitevitch and Luce (1998, 1999) among others.

Although numerous studies have demonstrated the importance of phonotactic probability in several language processes in native speakers of English and other languages (Vitevitch and Aljasser, 2021; Vitevitch and Luce, 2016), there is still much to learn about the use of phonotactic information in second language learners (Ulbrich et al., 2016).

To become proficient in another language, one must develop competence in many different types of knowledge about that language including syntactic, semantic, morphological, orthographic, and phonological information. According to the Common European Framework of Reference for Languages (Bailly et al., 2001; Section 5.2.1.4), among the phonological competencies that a speaker should have is "…a knowledge of, and skill in the perception and production of…the phonetic composition of words (syllable structure, the *sequence of phonemes* [emphasis added], word stress, word tones)" (Bailly et al., 2001, pp. 116–117).

Given the importance of phonotactic knowledge in native and second language learners, we assessed in the present study the phonotactic knowledge of adult learners of English by asking them to rate the word-likeness of English-like non-words that varied in phonotactic probability. This task was used previously to demonstrate that native speakers of English not only know which sounds and sequences of sounds are legal in English words, but also possess fine-grained knowledge about the frequency with which legal sounds and sequences of sounds appear in English words (Vitevitch et al., 1997).

In the present study, we examined the utility of the non-word rating task as an alternative or a complement to other tasks commonly used to assess proficiency in a second language. For example, reaction time-based tasks are often used to assess vocabulary and grammatical knowledge (Hui and Jia, 2024), but reaction time-based tasks may not exist or be suitable for assessing other types of linguistic knowledge in the second language learner. By using the non-word rating task we hoped to examine the implicit awareness of English learners to English phonotactic probability. If English learners are indeed sensitive to fine-grained phonotactic knowledge, then the non-word rating task may prove to be a short, simple but sensitive way to assess phonological knowledge in L2 learners, and may be useful for developing an alternative or complementary measure of language proficiency.

Previous work by Anisfeld and Gordon (1971) demonstrated that as little as 1 semester of experience with another language can modify phonotactic knowledge, so in the present study we compared participants (in Saudi Arabia and China) who had less experience with English (Beginners) to participants who had more experience with English (Advanced learners). We reasoned that ratings provided by Advanced learners of English would be more highly correlated with objective measures of phonotactic probability than ratings provided by Beginners, revealing subtle differences in the phonological competence of the two groups. By using English language learners from two different countries and with different ways of evaluating experience with English we hoped to also demonstrate that the assessment of phonotactic knowledge provided by the non-word rating task could be broadly used, and would not be tied to a specific language group, a particular method of assessing proficiency, etc. Given increasing concerns about a replication crisis in science (Udesky, 2025), our use of two different ways to measure experience with English also provides us an opportunity to internally replicate any effects that might be observed.

Methods

Participants

Learners of English were recruited from Qassim University in the Kingdom of Saudi Arabia (who were native speakers of Najdi Arabic) and from Kunming University of Science and Technology in the People's Republic of China (who were native speakers of Mandarin). Summary information about the participants is in Table 1. Data on gender/sex was not collected.

For the Arabic speakers, proficiency in English was determined by the semester of English classes that they were enrolled in at Qassim University. The Arabic-speaking participants who were currently enrolled in the first semester of English classes were categorized as Beginners, and participants who were currently enrolled in more advanced English classes were categorized as Advanced.

For the Mandarin speakers, proficiency in English was determined by their score on the College English Test (CET-Band 4), a standardized test created and administered by The National Advisory Committee on Teaching English Language to Non-English Majors in Higher Education under The Ministry of Education. The main goal of this standardized test is to measure how well students learn English at colleges in China (National

TABLE 1 Demographic information for the participants.

Native language	Arabic		Mandarin	
Proficiency level in English	Beginners	Advanced	Beginners	Advanced
Proficiency metric	1 semester (0.0)	5.2 semesters (1.8)	355 (37.7)	458 (43.2)
n	14	19	22	63
Age in years	18.6 (1.1)	21.9 (1.4)	19.5 (0.5)	19.9 (3.1)

Mean and standard deviation (in parentheses). The proficiency metric for the Mandarin speakers is the score on the CET-Band 4.

College English Testing Committee, 2006). Values on the 710-point scoring system of the CET-Band 4 may be used by Chinese colleges and universities to, among other things, guide admission decisions, and to determine if the English proficiency requirement for conferral of a bachelor's degree for non-English majors has been met. The values used to guide such decisions have changed over time as the overall level of English proficiency in China has increased. For example, the Ministry of Education previously recommended the value of 355 to satisfy the English proficiency requirement for conferral of a bachelor's degree for non-English majors, but more current guidelines recommend a value of 425 on the 710-point scoring system.

In the present study, the value of 400 on the CET-Band 4 represented a "natural break" in the distribution of sampled scores. The Chinese-speaking participants in the present study who scored below 400 on the CET-Band 4 were categorized as Beginners, and participants who scored 400 or above were categorized as Advanced.

We recognize that categorizing learners into Beginners and Advanced proficiency groups may appear to be a course-grained distinction. Recall, however, the work of Anisfeld and Gordon (1971), which showed that as little as 1 semester of experience with another language can influence phonotactic knowledge, justifying the broad distinction used in the present study. Further, our use of two different metrics (i.e., number of English classes and score on a standardized test) allows us to directly address concerns of reliability, replicability, and generalizability by demonstrating that any effects we may observe are not restricted to a specific measure of experience, metric of proficiency, etc.

This project was approved by the Institutional Review Boards at the University of Kansas, Qassim University, and Kunming University of Science and Technology. All participants provided informed consent to participate in this study.

Stimuli

The stimuli were the 60 consonant-vowel-consonant, monosyllabic non-words (30 with high phonotactic probability, 30 with low phonotactic probability) previously used in Vitevitch and Donoso (2012). The phonotactic probabilities for the stimuli were calculated using the Phonotactic Probability Calculator (Vitevitch and Luce, 2004). As previously reported in Vitevitch and Donoso (2012), the high phonotactic probability non-words, such as /dʌp/, had sum of segments (*mean* = 0.167, *SEM* = 0.004) and sum of sequences of segments (*mean* = 0.086, *SEM* = 0.004) that were significantly higher [$F_{(1,58)} = 193.57$, $p < 0.0001$] than the sum of segments (*mean* = 0.008, *SEM* = 0.001) and sum of sequences of segments (*mean* = 0.001, *SEM* = 0.0001) for the low phonotactic probability non-words, such as /ʧɜ~ʃ/. An equal number of non-words in each condition contained the same initial consonants (3 non-words each started with /b/, /d/, /f/, /g/, /ʤ/, /m/, /n/, /p/, /r/, /t/).

As described in Vitevitch and Luce (2004), position specific segment probability (used to calculate the sum of segments) was calculated by searching a computer readable dictionary for words (regardless of word length) that contained a given segment in a given position in a word. The frequency values of those words were summed together and then divided by the total frequency of all the words in the dictionary that had a segment in that position to provide an estimate of position specific segment probability. Similarly, the probability of the sequences of segments (specifically, position specific biphones were used to obtain the sum of sequences of segments) was calculated by searching a computer readable dictionary for words (regardless of word length) that contained all instances in which a sequence of two phonemes occurred together in specific adjacent positions in a word. The frequency of occurrence for the words in which the (position-specific) biphoneme sequences were found were summed and then divided by the summed of the frequency counts for all words in the dictionary that contained phonemes in those two adjacent positions. See Mayer et al. (2025) and others for alternative ways to compute phonotactic probability.

Procedure

A non-word rating task as used in Vitevitch et al. (1997; see also Vitevitch, 2025) was used in the present study. A typical trial proceeded as follows: A scale from 1, labeled "Good English Word," to 7, labeled "Bad English Word," appeared on the computer screen. Participants were presented auditorily with one of the stimulus items at a comfortable listening level, and instructed to select one of the labeled buttons 1 through 7 as quickly as possible. After the response was recorded, the next trial began. Qualtrics software was used to randomly present the auditory stimuli (in mp3 format) and to collect responses.

Results

The mean rating for each of the 60 stimulus items was computed for the Beginners and Advanced Arabic learners of English, and for the Beginners and Advanced Mandarin learners of

TABLE 2 Pearson *r* values for the learners of English.

Native language	Arabic		Mandarin	
	Beginners	Advanced	Beginners	Advanced
Sum of segments	0.47	0.63	0.52	0.65
Sum of sequences of segments	0.45	0.51	0.49	0.50

All r-values were statistically significant at $p < 0.0001$.

English. For each of the 4 groups of learners, a Pearson correlation coefficient was calculated (JASP Team, 2022) to assess the relationship of the mean subjective rating provided by the learners and the two objective measures of phonotactic probability—sum of the segments and sum of the sequences of segments—obtained from the Phonotactic Probability Calculator (Vitevitch and Luce, 2004). The data that support the findings of this study are available on request from the corresponding author, and are summarized in Table 2. For comparison, the subjective ratings to the same items provided by 19 native English speakers were significantly correlated with the sum of the segments at $r = 0.55$, and with the sum of the sequences of segments at $r = 0.56$, as reported in Vitevitch and Donoso (2012).

All the correlations reported in Table 2 indicate a statistically significant relationship between the subjective ratings and the objective measures of phonotactic probability (i.e., all p's < 0.0001). Further, the coefficient of determination (i.e., r^2) indicates that the size of the observed effects ranged from medium to large. For reference, commonly accepted values of $r^2 = 0.01$ is considered a small effect, $r^2 = 0.09$ is considered a medium effect, and $r^2 = 0.25$ is considered a large effect.

Analysis using Fisher r-to-z transformation to test for a significant difference between any two correlation coefficients reported in Table 2 (including the values from the native English speakers reported in Vitevitch and Donoso, 2012) showed no significant difference between any pair of correlation coefficients. That is, the correlations obtained for the Beginners was comparable to the correlations obtained for the Advanced learners. Similarly, the correlations obtained for the native English speakers (as reported in Vitevitch and Donoso, 2012) were also comparable to the correlations obtained for the Beginners and Advanced learners.

Turning from the relationship between the subjective word-likeness ratings and the objective measures of phonotactic probability, we now consider the relationship between the subjective word-likeness ratings of one group and the subjective word-likeness ratings of another group. For Arabic learners of English, the Pearson correlation coefficient shows a significant and large relationship between the subjective ratings of the Beginners and the subjective ratings of the Advanced learners [$r_{(60)} = 0.78$, $p < 0.0001$, $r^2 = 0.61$ a large effect]. A significant and large relationship was also found between the subjective ratings of the native English speakers (as reported in Vitevitch and Donoso, 2012) and the subjective ratings of the Beginners [$r_{(60)} = 0.55$, $p < 0.0001$, $r^2 = 0.30$ a large effect]. Similarly, there was a significant and large relationship between the subjective ratings of the native English speakers (as reported in Vitevitch and Donoso, 2012) and the subjective ratings of the Advanced learners [$r_{(60)} = 0.51$, $p < 0.0001$, $r^2 = 0.26$ a large effect].

Analysis using Fisher r-to-z transformation to test for a significant difference between two correlation coefficients shows that the correlation between the subjective ratings of the Arabic Beginner and Advanced learners ($r = 0.78$) was significantly greater than the correlation between the subjective ratings of Beginners and native speakers [$r = 0.55$; $z_{(60)} = 2.28$, $p < 0.05$]. Similarly, the correlation between the subjective ratings of Beginner and Advanced learners ($r = 0.78$) was significantly greater than the correlation between the subjective ratings of Advanced learners and native speakers [$r = 0.51$; $z_{(60)} = 2.58$, $p < 0.01$]. No other comparison was significant different.

A similar analysis of the subjective word-likeness ratings from Chinese learners of English was also performed. For Chinese learners of English, the Pearson correlation coefficient shows a significant and large relationship between the subjective ratings of the Beginners and the subjective ratings of the Advanced learners [$r_{(60)} = 0.82$, $p < 0.0001$, $r^2 = 0.67$ a large effect]. A significant and large relationship was also found between the subjective ratings of the native English speakers (as reported in Vitevitch and Donoso, 2012) and the subjective ratings of the Beginners [$r_{(60)} = 0.56$, $p < 0.0001$, $r^2 = 0.31$ a large effect]. Similarly, there was a significant and large relationship between the subjective ratings of the native English speakers (as reported in Vitevitch and Donoso, 2012) and the subjective ratings of the Advanced learners [$r_{(60)} = 0.60$, $p < 0.0001$, $r^2 = 0.36$ a large effect].

Analysis using Fisher r-to-z transformation to test for a significant difference between two correlation coefficients shows that the correlation between the subjective ratings of the Chinese Beginner and Advanced learners ($r = 0.82$) was significantly greater than the correlation between the subjective ratings of Beginners and native speakers [$r = 0.56$; $z_{(60)} = 2.64$, $p < 0.01$]. Similarly, the correlation between the subjective ratings of Beginner and Advanced learners ($r = 0.82$) was significantly greater than the correlation between the subjective ratings of Advanced learners and native speakers [$r = 0.60$; $z_{(60)} = 2.32$, $p < 0.05$]. No other comparison was significant different.

To further examine the subjective word-likeness ratings at a group level we used ANOVA to compare the ratings from the Beginners, Advanced learners, and native speakers (as reported in Vitevitch and Donoso, 2012) for the Arabic and Chinese learners of English. For simplicity these analyses do not include phonotactic probability as a factor (however, the results are qualitatively similar if both proficiency level and phonotactic probability are included in the analysis). For the Arabic learners of English, a significant difference was found in the subjective word-likeness ratings for the Beginners, Advanced learners, and native speakers [$F_{(2,177)} = 2.52$, $p < 0.01$, $\omega^2 = 0.02$ a small effect]. *Post-hoc* comparisons (with Tukey corrections) show that the Beginners (*mean* = 3.75; *sd* =

0.97) tended to rate the non-words higher (though not significantly so) than the native speakers (*mean* = 3.53; *sd* = 0.76). Similarly, the Advanced learners (*mean* = 3.90; *sd* = 0.98) tended to rate the non-words higher (though not significantly so) than the native speakers (*mean* = 3.53; *sd* = 0.76). There was also no statistical difference between the subjective word-likeness ratings for the Beginners and the Advanced learners.

For the Chinese learners of English, a significant difference was also found in the subjective word-likeness ratings for the Beginners, Advanced learners, and native speakers [$F_{(2,177)} = 77.93$, $p < 0.0001$, $\omega^2 = 0.46$ a medium effect]. *Post-hoc* comparisons (with Tukey corrections) show that the Beginners (*mean* = 5.01; *sd* = 0.72) rated the non-words significantly higher [$t_{(177)} = 10.49$, $p < 0.0001$] than the native speakers (*mean* = 3.53; *sd* = 0.76). Similarly, the Advanced learners (*mean* = 5.10; *sd* = 0.84) rated the non-words significantly higher [$t_{(177)} = 11.11$, $p < 0.0001$] than the native speakers (*mean* = 3.53; *sd* = 0.76). There was no statistical difference between the subjective word-likeness ratings for the Beginners and the Advanced learners.

Discussion

In the present study we asked native speakers of Arabic and Chinese who were learning English to provide word-likeness ratings of English-like non-words. This task has been used previously with native speakers of English to demonstrate their sensitivity to phonotactic probability in English (e.g., Vitevitch et al., 1997). Our goal in the present study was to determine if the same task could be used with learners of English as a short, simple but potentially sensitive way to assess their implicit knowledge of English phonotactics, and as an alternative means to assess English proficiency.

We examined two groups of learners of English who differed in their native language (Arabic or Chinese). These groups also differed in how "proficiency" was assessed, using either scores on a standardized test (as in China) or the number of English classes taken (as in Saudi Arabia). By assessing the word-likeness ratings of these two groups, we hoped to demonstrate that word-likeness ratings might be useful in a variety of settings, situations, populations, etc.

Our initial prediction was that the subjective word-likeness ratings of English-like non-words provided by Advanced learners of English would be more highly correlated with two objective measures of phonotactic probability (i.e., the sum of segments, and the sum of sequences of segments) than the subjective word-likeness ratings provided by Beginners learning English. The correlations between the subjective ratings and the objective measures of phonotactic probability showed medium to large effects (as measured by r^2) for all groups, indicating that the non-word rating task does assess sensitivity to phonotactic probability in learners of English, not just in native speakers of English (e.g., Vitevitch et al., 1997). Contrary to our prediction, however, no difference was observed between the correlations for the Beginners and Advanced learners in Saudi Arabia or in China (as determined by Fisher *r*-to-*z* transformation tests).

Given that Anisfeld and Gordon (1971) found that as little as one semester of experience with another language can affect phonotactic knowledge (not only of the language being learned but of the native language as well), we were perhaps too optimistic in predicting that the non-word rating task would be sensitive enough to discriminate between Beginners and Advanced learners of English. We further reasoned that the goal of learning another language is to sound like a native speaker of that language, not to sound better than the person sitting next to you in the language class. Therefore, we performed additional analyses comparing in various ways the subjective ratings of Beginner and Advanced learners of English to native speakers of English (as reported in Vitevitch and Donoso, 2012).

In several analyses we found significant correlations between the subjective ratings of word-likeness for the Beginners and Advanced learners in both the Arabic and Chinese learners of English. The subjective ratings of word-likeness for the Beginners and Advanced learners in both the Arabic and Chinese learners of English also correlated with the subjective ratings of word-likeness from the native speakers of English (as reported in Vitevitch and Donoso, 2012). Using Fisher *r*-to-*z* transformation tests we further found that the correlation of subjective ratings between the Beginner and Advanced learners was significantly larger than the correlation of subjective ratings between the Beginners and native speakers, and between the Advanced learners and native speakers. That is, as one might expect given the findings of Anisfeld and Gordon (1971), the Beginner and Advanced learners are more similar to each other in their knowledge of phonotactics than either group is to native speakers of English. This also suggests that rather than using the non-word rating task to discriminate between Beginner and Advanced learners, it might be more fruitful to use the task to see how similar Beginner or Advanced learners are to native speakers.

We then used ANOVA to examine group differences in the subjective ratings of word-likeness for the Beginner and Advanced learners, and native speakers of English (as reported in Vitevitch and Donoso, 2012). In this analysis we found that Beginner and Advanced learners provided higher word-likeness ratings than the native speakers of English (this pattern was statistically significant only for the Chinese learners of English). There may be many reasons for such a pattern, but we present below two hypothesis that will require future studies to test.

One possible reason that Beginner and Advanced learners provided higher word-likeness ratings than native speakers of English is that native speakers of English may not rely solely on phonotactic knowledge to make their ratings. Perhaps, the non-word sounds similar to a real word, allowing semantic information to influence the rating (Levy et al., 2021). Or perhaps the non-word just sounds "funny," which might influence the word-likeness rating in some way (see Study 3 in Vitevitch, 2025 and Westbury and Hollis, 2019).

Another possibility is that even Advanced learners of English may not have been exposed to as many and as wide a variety of English words as a native speaker of English. Because learners of English may not have encountered a large number of rarely occurring, arcane, or otherwise unusual English words, every

English-like non-word may sound pretty good to them. Only after being exposed to several rarely occurring, arcane, or otherwise unusual English words might a learner of English acquire a better sense of how bad an English word could sound, enabling the learner to adjust their word-likeness ratings to be more consistent with those of a native English speaker (see Martinez and Vitevitch, 2024, for an example of how vocabulary size might affect the creative use of language by learners of a language). The present data do not allow us to distinguish between these two possibilities, or to rule out other factors that may account for the higher word-likeness ratings observed for language learners. We await future research to address this issue.

Previous studies demonstrated in native speakers that knowledge of the segments and sequences of segments that form a word (i.e., phonotactic knowledge) influences several cognitive processes including word learning, word segmentation, word recognition, and word production (Vitevitch and Aljasser, 2021). It stands to reason that phonotactic knowledge might also influence the same cognitive processes (e.g., word learning, word segmentation, word recognition, and word production) in second language learners.

Indeed, the work of Kivistö-de Souza (2017) shows that the phonotactic knowledge of native speakers of Brazilian Portuguese learning English as measured in a lexical decision task correlated with their pronunciation of English as measured in the Foreign Accent Rating task. That is, learners of English who showed greater sensitivity to phonotactic information in a perceptual task tended to produce better English pronunciations as determined by ratings provided by a panel of judges in the Foreign Accent Rating task. This finding prompted Kivistö-de Souza to "…suggest that phonotactics should be taught in foreign language classrooms since increasing learners' awareness might be beneficial for the accuracy of their L2 pronunciation" (Kivistö-de Souza, 2017, p. 185). The relationship between phonotactic knowledge and native-like pronunciation observed by Kivistö-de Souza (2017) further suggests that assessments of phonotactic knowledge could be useful as additional or alternative behavioral markers of language proficiency.

The results of the present study suggest that (with some refinement) a task asking language learners to subjectively rate the word-likeness of English-like non-words has potential to be used as an additional or alternative behavioral marker of language proficiency. The value of a word-likeness rating task increases when one considers the demands and resources required by current assessments of language proficiency.

Consider for example, the Foreign Accent Rating task as used by Kivistö-de Souza (2017) and others. In the typical Foreign Accent Rating task, a panel of native (or non-native) judges rate the degree of foreign accent in multiple speech samples (often sentence length utterances) produced by language learners (e.g., Bongaerts et al., 1997; Flege, 1988). Although there is often high agreement among the judges, the typical panel of judges contains more than one person. The requirement of having multiple judges could be a limiting factor for small or underfunded language learning programs that do not have ready access to several native/proficient speakers to serve on a panel for the Foreign Accent Rating task. Further, increasing the number of students producing sentence-length utterances to-be-rated increases burden (e.g., time and effort) on the judges. Furthermore, L2 learners may produce statistically reliable distinctions between sounds that even trained listeners may not perceive, referred to as covert contrasts (e.g., Eckman et al., 2014; Tyler et al., 1993), increasing the motivation to use additional or alternative means to assess the knowledge of the speaker.

In contrast, in the word-likeness rating task used in the present study, subjective ratings to 60 non-words were provided by the language learners themselves in a session that lasted approximately 15 min. The number of language learners engaged in the rating task did not affect the duration of the task, nor increase burden on the assessors. Further, the subjective ratings provided by the language learners were compared to objective measures of phonotactic probability that were obtained from a freely available online Phonotactic Probability Calculator (e.g., Aljasser and Vitevitch, 2018; Vitevitch and Luce, 2004) by a single assessor (not judgments provided by multiple assessors). Thus, the relationship between subjective word-likeness ratings and objective measures of phonotactic probability observed in the present study using a simple rating task performed by the language learners themselves may provide language instructors at all levels with an efficient, low-cost, and low-effort way that is not affected by perceptual limitations (i.e., covert contrasts) to assess and track the development of one aspect of language proficiency in second language learners. Given the number of language processes that phonotactic information influences (e.g., Vitevitch and Aljasser, 2021), the present finding, though preliminary in nature, could have broad implications for efficiently assessing numerous aspects of proficiency in language learners with minimal effort for a single instructor.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Institutional Review Boards at the University of Kansas, Qassim University, and Kunming University of Science and Technology.

Author contributions

FA: Conceptualization, Investigation, Methodology, Software, Supervision, Writing – original draft. YW: Conceptualization, Investigation, Software, Writing – original draft, Writing – review & editing. MV: Conceptualization, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing.

Funding

The authors declare that financial support was received for the research reported in this article. APC was funded by an agreement with the University of Kansas Libraries.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

References

Al-jasser, F. (2008). The effect of teaching English phonotactics on the lexical segmentation of English as a foreign language. *System* 36, 94–106. doi: 10.1016/j.system.2007.12.002

Aljasser, F., and Vitevitch, M. S. (2018). A web-based interface to calculate phonotactic probability for words and nonwords in Modern Standard Arabic. *Behav. Res. Methods* 50, 313–322. doi: 10.3758/s13428-017-0872-z

Anisfeld, M., and Gordon, M. (1971). An effect of one German-language course on English. *Lang. Speech* 14, 289–292. doi: 10.1177/002383097101400309

Bailly, S., Devitt, S., Gremmo, M.-J., Heyworth, F., Hopkins, A., Jones, B., et al. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. A Guide for Users*. Strasbourg: Language Policy Division.

Bongaerts, T., van Summeren, C., Planken, B., and Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Stud. Second Lang. Acquisit.* 19, 447–465. doi: 10.1017/S0272263197004026

Dal Ben, R., Souza, D. H., and Hay, J. F. (2021). When statistics collide: the use of transitional and phonotactic probability cues to word boundaries. *Memory Cogn.* 49, 1300–1310. doi: 10.3758/s13421-021-01163-4

Eckman, F., Iverson, G., and Song, J. Y. (2014). "Covert contrast in the acquisition of second language phonology," in *Perspectives on Phonological Theory and Development: In Honor of Daniel A. Dinnsen*, eds. A. W. Farris-Trimble and J. A. Barlow (Philadelphia, PA: John Benjamins Publishing Company), 25–48. doi: 10.1075/lald.56.05eck

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *J. Acoust. Soc. Am.* 84, 70–79. doi: 10.1121/1.396876

Goldrick, M., and Larson, M. (2008). Phonotactic probability influences speech production. *Cognition* 107, 1155–1164. doi: 10.1016/j.cognition.2007.11.009

Gonzalez-Gomez, N., Poltrock, S., and Nazzi, T. (2013). A "bat" is easier to learn than a "tab": effects of relative phonotactic frequency on infant word learning. *PLoS ONE* 8:e59601. doi: 10.1371/journal.pone.0059601

Graf Estes, K., Edwards, J., and Saffran, J. R. (2011). Phonotactic constraints on infant word learning. *Infancy* 16, 180–197. doi: 10.1111/j.1532-7078.2010.00046.x

Hoover, J. R., Storkel, H. L., and Hogan, T. P. (2010). A cross-sectional comparison of the effects of phonotactic probability and neighborhood density on word learning by preschool children. *J. Mem. Lang.* 63, 100–116. doi: 10.1016/j.jml.2010.02.003

Hui, B., and Jia, R. (2024). Reflecting on the use of response times to index linguistic knowledge in SLA. *Annu. Rev. Appl. Linguist.* 1–11. doi: 10.1017/S0267190524000047

JASP Team (2022). *JASP, Version 0.18.3; Computer Software*. Amsterdam: JASP Team.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y., and Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *J. Mem. Lang.* 32, 402–420. doi: 10.1006/jmla.1993.1022

Jusczyk, P. W., Luce, P. A., and Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *J. Mem. Lang.* 33, 630–645. doi: 10.1006/jmla.1994.1030

Kelley, M. C., and Tucker, B. V. (2022). The recognition of spoken pseudowords. *Lang. Cogn. Neurosci.* 37, 1169–1190. doi: 10.1080/23273798.2022.2053729

Kivistö-de Souza, H. (2017). The relationship between phonotactic awareness and pronunciation in adult second language learners. *Rev. Brasil. Linguísta Aplicada* 17, 185–214. doi: 10.1590/1984-6398201610850

Levy, O., Kenett, Y. N., Oxenberg, O., Castro, N., DeDeyne, S., Vitevitch, M. S., et al. (2021). Unveiling the nature of interaction between semantics and phonology in lexical access based on multilayer networks. *Sci. Rep.* 11:14479. doi: 10.1038/s41598-021-93925-y

MacRoy-Higgins, M., and Dalton, K. P. (2015). The influence of phonotactic probability on nonword repetition and fast mapping in 3-year-olds with a history of expressive language delay. *J. Speech Lang. Hearing Res.* 58, 1773–1779. doi: 10.1044/2015_JSLHR-L-15-0079

Martinez, A. E., and Vitevitch, M. S. (2024). A cognitive network analysis of semantic associates in monolingual English speakers and learners of Kaqchikel. *Biling. Lang. Cogn.* 27, 1046–1052. doi: 10.1017/S1366728924000762

Mattys, S. L., and Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition* 78, 91–121. doi: 10.1016/S0010-0277(00)00109-8

Mattys, S. L., Jusczyk, P. W., Luce, P. A., and Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cogn. Psychol.* 38, 465–494. doi: 10.1006/cogp.1999.0721

Mayer, C., Kondur, A., and Sundara, M. (2025). The UCI phonotactic calculator: an online tool for computing phonotactic metrics. *Behav. Res. Methods* 57:218. doi: 10.3758/s13428-025-02725-z

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *J. Mem. Lang.* 39, 21–46. doi: 10.1006/jmla.1998.2568

Mersad, K., and Nazzi, T. (2011). Transitional probabilities and positional frequency phonotactics in a hierarchical model of speech segmentation. *Memory Cogn.* 39, 1085–1093. doi: 10.3758/s13421-011-0074-3

National College English Testing Committee (2006). *Syllabus for College English Test Band 4*. Shanghai: Shanghai Foreign Language Education Press.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928. doi: 10.1126/science.274.5294.1926

Storkel, H. L., Armbruster, J., and Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *J. Speech Lang. Hearing Res.* 49, 1175–1192. doi: 10.1044/1092-4388(2006/085)

Storkel, H. L., and Maekawa, J. (2005). A comparison of homonym and novel word learning: the role of phonotactic probability and word frequency. *J. Child Lang.* 32, 827–853. doi: 10.1017/S0305000905007099

Storkel, H. L., and Rogers, M. A. (2000). The effect of probabilistic phonotactics on lexical acquisition. *Clin. Linguist. Phonetics* 14, 407–425. doi: 10.1080/026992000415859

Sundara, M., Zhou, Z. L., Breiss, C., Katsuda, H., and Steffman, J. (2022). Infants' developing sensitivity to native language phonotactics: a meta-analysis. *Cognition* 221:104993. doi: 10.1016/j.cognition.2021.104993

Tyler, A. A., Figurski, G. R., and Langdale, T. (1993). Relationships between acoustically determined knowledge of stop place and voicing contrasts and phonological treatment progress. *J. Speech Hear. Res.* 36, 746–759. doi: 10.1044/jshr.3604.746

Udesky, L. (2025). 'Publish or perish' culture blamed for reproducibility crisis. *Nature* 10.1038/d41586-024-04253-w. doi: 10.1038/d41586-024-04253-w

Ulbrich, C., Alday, P. M., Knaus, J., Orzechowska, P., and Wiese, R. (2016). The role of phonotactic principles in language processing. *Lang. Cogn. Neurosci.* 31, 662–682, doi: 10.1080/23273798.2015.1136427

Vitevitch, M. S. (2025). Examining Chat GPT with nonwords and machine psycholinguistic techniques. *PLoS ONE* 20: e0325612. doi: 10.1371/journal.pone.0325612

Vitevitch, M. S., and Aljasser, F. M. (2021). "Phonotactics in spoken-word recognition," in *The Handbook of Speech Perception*, eds. J. S. Pardo, L. C. Nygaard, R. E. Remez, and D. B. Pisoni (Wiley-Blackwell), 286–308. doi: 10.1002/9781119184096.ch11

Vitevitch, M. S., and Donoso, A. (2012). Phonotactic probability of brand names: I'd buy that! *Psychol. Res.* 76, 693–698. doi: 10.1007/s00426-011-0374-z

Vitevitch, M. S., and Luce, P. (2016). Phonological neighborhood effects in spoken word perception and production. *Annu. Rev. Linguist.* 2, 75–94. doi: 10.1146/annurev-linguistics-030514-124832

Vitevitch, M. S., and Luce, P. A. (1998). When words compete: levels of processing in spoken word perception. *Psychol. Sci.* 9, 325–329.

Vitevitch, M. S., and Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *J. Mem. Lang.* 40, 374–408.

Vitevitch, M. S., and Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behav. Res. Methods Instruments Comput.* 36, 481–487. doi: 10.3758/BF03195594

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., and Kemmerer, D. (1997). Phonotactics and syllable stress: implications for the processing of spoken nonsense words. *Lang. Speech* 40, 47–62. doi: 10.1177/002383099704000103

Weber, A., and Cutler, A. (2006). First-language phonotactics in second-language listening. *J. Acoust. Soc. Am.* 119, 597–607. doi: 10.1121/1.2141003

Westbury, C., and Hollis, G. (2019). Wriggly, squiffy, lummox, and boobs: what makes some words funny? *J. Exp. Psychol. General* 148, 97–123. doi: 10.1037/xge0000467

Zamuner, T. S. (2009). Phonotactic probabilities at the onset of language development: speech production and word position. *J. Speech Lang. Hearing Res.* 52, 49–60. doi: 10.1044/1092-4388(2008/07-0138)