

# PRONUNCIATION EVALUATION CRITERIA FOR EFL LEARNERS

**Hana Vančová**

*Department of English Language and Literature, Faculty of Education,  
Trnava University in Trnava (Slovakia)*

## Abstract

Pronunciation is one of the competencies foreign language learners of English are implicitly or explicitly judged for in classroom context as well as real-life communication. At the same time, both teachers and learners express concerns concerning this competence, as relatively little attention is being paid to pronunciation issues. While accuracy was desired in the past, comprehensibility is preferred as the goal of pronunciation instruction in recent years. Mistakes and errors the speakers make vary across the language background of speakers and can be manifested at segmental and suprasegmental levels; however, familiarity with the topic discussed may help overcome many obstacles the foreign language speakers may make. Pronunciation may be judged by human raters as well as automatically by specialized software. The presented study aims to current practices presented in research papers published in the past ten years. The results suggest different criteria applied to pronunciation evaluation.

The paper presents partial research outcomes of the projects KEGA 019TTU-4/2021 Introducing new digital tools into teaching and research within transdisciplinary philological study programmes and 7/TU/2021 Pronunciation mistakes of pre-service teachers of English.

**Keywords:** *EFL, pronunciation, evaluation, comprehensibility, accuracy.*

---

## 1. Introduction

Evaluation in education is defined in the broader context as assessing “merit and worth” (Byram & Hu, 2017, p.234). Evaluation, therefore, considers the evidence provided in the process, which should methodologically justify the individual steps taken in the process of teaching and evaluation. Kizlik (2012, p. 2) suggests evaluation allows interpretation of the collected data within a “situation”, i.e., material and other conditions in which evaluation occurs. This justification is possible due to the multidisciplinary nature of evaluation, as it considers models and approaches from different disciplines. More specifically, educators (e.g., Byram & Hu, 2017, Scanlan, 2012) distinguish between formative assessment, i.e., strengthening the evaluated subject, and summative evaluation, i.e., the result of the subject achieved after delivering a teaching programme.

According to the Common European Framework of Reference for Languages: learning, teaching, assessment; Council of Europe, Companion Volume (CEFR; 2018), a key document shaping language policies in Europe, spoken communication relies on pronunciation, which is one of the competencies a foreign language learner must master to be able to participate in a broader social discourse. As a result, the document focuses on various degrees of comprehensibility and intelligibility according to the proficiency levels foreign learners achieve. Through language, foreign language learners express themselves and adapt to their language community using acoustic forms typical and comprehensible for a particular setting. Therefore, pronunciation is one of the first features of language a communication partner perceives and forms their attitude towards the speaker (Hendriks, van Meurs & Usmany, 2021).

The dominant approaches have shaped the evaluation of non-native pronunciation in foreign language teaching. While audiolingualism emphasizes pronunciation and various forms of its training, in the currently predominant communicative language teaching, learners are expected to achieve good pronunciation by exposure and attention to other layers of language (Murcia, Brinton, & Goodwin, 1996). The most current perspective on pronunciation teaching emphasizes comprehensibility and intelligibility (Levis, 2005; Vančová, 2019). Under the approaches mentioned above, two general goals and the following evaluation criteria for pronunciation are considered. The first is the ability to imitate the native-like pronunciation of non-native learners, which aims for accuracy. Accuracy is evaluated by comparing the model’s and learner’s pronunciation, highlighting the deviations. The degree of proximity of sounds of a native pronunciation and the deviation of a non-native speaker is referred to as an

accentedness and can be judged by human raters or digital tools (Levis, 2007; Pokrivčáková, 2015). The second is the speaker's ability to be understood by listeners judged from the listener's perspective. This type of pronunciation skill is referred to as intelligibility or comprehensibility (Vančová, 2021). While intelligibility refers to the speaker's ability to be understood, the latter refers to the listeners' amount of effort to understand the speaker. Intelligibility and comprehensibility, contrary to accuracy, are predominantly judged by human raters who assess these aspects of non-native speech by transcription or comprehensibility evaluation sheets. Speech sounds are assigned a high or low functional load which hinders or promotes comprehensibility (Munro, 2010). However, judging these pronunciation aspects may be subjective, as human raters' experience with judging pronunciation may affect the results of their evaluation (Isaacs & Thompson, 2013), as well as external factors, e.g., raters' tiredness or background noise (Isaacs & Harding, 2017; Sheppard, Elliott & Baese-Berk, 2017).

Linguistically homogeneous classes can address the same type of mistakes in terms of accuracy. The possible deviations in pronunciation of foreign learners can be typically grouped according to the influence of the mother tongue (Kelly, 2000). These challenges then present the aim of the pronunciation instruction in accuracy. To know the needs their learners should address, teachers identify the typical mistakes their linguistically homogeneous groups make. As far as the pronunciation requirements and proficiency levels are concerned, the goals for each level of foreign language learners are not clearly defined. Thus, many teachers and learners are predominantly guided by course books, which present segmentals and suprasegmentals across all proficiency levels, and the purpose they use the English language (non-professional or expert use, Vančová, 2020, 2021).

## 2. Design

The presented study follows the qualitative approach, which aims to “understanding behaviors of values, beliefs and assumptions” (Choy, 2014, p. 101). The emphasis is placed on content analysis of recent research studies published in peer-reviewed academic journals and proceedings (since 2012).

## 3. Objectives

The presented study aims to answer the following questions:

1. *What criteria do the authors apply for pronunciation evaluation?*
2. *In recent research studies dealing with pronunciation improvement, do the researchers use human raters or digital tools for English pronunciation evaluation?*

## 4. Methods

To answer the research questions, a method of content analysis was selected. The content analysis focused on the data presented in research studies presented in reviewed academic journals or conference proceedings and published since 2012 and available in academic databases. The research studies were searched for in the following databases: *Google Scholar, Eric, Taylor and Francis, Web of Science Core Collection, Wiley Online Library, SAGE and Science Direct* using the keywords “pronunciation, evaluation, English, rubrics, scale”. The search results were then checked for appropriateness (presenting original research data of pronunciation evaluation by human raters or digital tools). After that, ten studies from the following journals and conference proceedings were selected: *Asian EFL Journal Research Articles, Proceedings of the International Symposium on Automatic Detection of Errors in Pronunciation Training June 6 – 8, 2012 KTH, Stockholm, Sweden, Special issue Making Connections: Studies of Language and Literature Education (Special issue in honor of Gert Rijlaarsdam) (2018), Language Teaching Research, Language Learning & Technology, Computer Assisted Language Learning, System and IAFOR Journal of Education.*

## 5. Results

The selected articles were analyzed for the overall conditions in the evaluation process and the identified criteria for pronunciation evaluation in the context. The overview is presented in Table 1 below:

Table 1. Overview of the analyzed studies.

Study	Participants and raters	Instruction and/or data collection tool	Pronunciation focus
Zahra et al. (2012)	9 EFL learners (speakers) and 12 native-speaking raters	decision tree and speech recognition technology	obvious, minor, no errors (phone and syllable level)
Luo (2014)	55 Taiwanese English majors and two human raters, peer-review	computer-assisted pronunciation training technique	4-point holistic scale (native-like to incomprehensible, segments to sentences)
Fouz-González (2017)	121 Spanish learners of English, 2 non-native judges	Twitter-based instruction	mispronunciation of words (phonemes, lexical stress)
Gluhareva & Prieto (2017)	20 Catalan undergraduates (18-30) and 5 native speakers of American English	recorded speech (beat and non-beat)	overall comprehensibility (rhythm)
Zoghbor (2017)	50 L1 Arabic learners of English and 18 non-native and native English-speaking raters	Lingua Franca Core features, semi-structured interviews	pronunciation features promoting intelligibility of non-native learners of English
Koet & van den Bergh (2018)	20 Dutch upper-intermediate to advanced learners and 126 listeners	recorded speech samples (non-communicative)	comprehensibility, aesthetic quality, intonation, standardness
Al-Ahdal (2020)	32 Saudi EFL learners	podcasts (TED Talks), in-class discussion	not specified
Fouz-González (2020)	52 Spanish learners of English	English File Pronunciation app	fossilized vowels, alveolar fricatives
Moxon (2021)	105 Thai undergraduate students	SpeechAce	phonetic accuracy
Suzukida & Saito (2022)	40 Japanese learners and native-speaking raters	monologue speaking task, IELTS Pronunciation Scale	segmentals and suprasegmentals, word stress and intonation

The earlier stages of pronunciation evaluation using technology compared human evaluation and technology evaluation. Zahra et al. (2012) compared the evaluation of human and software ratings of pronunciation of ELF speakers of different linguistic backgrounds (Syrian, Dutch, Portugal, Indonesian, Pakistani, Hungarian, Nigerian and Irish). The results were not consistent (false positivity and negativity) due to two possible reasons – low quality of the system or little strictness of human raters. The study was not carried out for educational purposes but system testing. Similarly, Luo (2014) compared the effectivity of the experimental group's computer-assisted pronunciation training (CAPT) with the results achieved by the in-class only trained participants in the control group. Students provided peer feedback by commenting on a discussion board. Pronunciation was evaluated by 4-point rating rubrics, which contrasted native-like and intelligible pronunciation on different levels (phoneme, clusters, words, sentences). The peer-review of multiple raters ensured various mistakes were noted as each reviewer tended to concentrate on other types of mistakes. As a result, this type of peer feedback appeared to be sufficiently reliable. Finally, Moxon (2021) identified a statistically significant correlation between the frequency of practice and pronunciation accuracy in Thai undergraduate learners of English using speech recognition software SpeechAce.

Koet & van den Bergen (2018) used semantic differential to compare evaluation criteria of Dutch and English raters for describing the Dutch learners' pronunciation in English on 7-point scales (positive attributes: pleasant, cultured, beautiful, polished, no accent, standard, melodious, expressive, intelligible, precise, distinguished). In addition, listeners focused on speakers' aesthetic quality, intonation, comprehensibility and standardness. The findings suggest (concerning study limitations) that non-native listeners were unreliable in evaluating the aesthetic quality and intonation; however, the raters appeared to be reliable in judging comprehensibility. Such results, therefore, limit the use of non-native listeners as examiners.

Gluhareva & Prieto (2017) aimed to use gestures to improve the participants' rhythm in speech and thus let native speaking raters evaluate the overall comprehensibility, stress and intonation. The authors also considered that intelligible speech could be accented, which means that this focus of the evaluation is more inclusive than the evaluation of native-like, non-accented speech. In addition, the

samples were presented to raters in pairs (with and without the beat) to increase the raters' sensitivity during evaluation.

Regarding accuracy, Zoghbor (2017) used Jenkins' Lingua Franca Core features promoting the intelligibility of non-native learners of English as the basis for pronunciation training. The native as well as non-native speakers of English judged the performances and indicated all instances of unintelligible utterance. Similarly, Fouz-González (2020) concentrated on the production and perception of English vowels /æ, ɑ:, ʌ, ə/ and alveolar consonants /s, z/ B2 Spanish learners of English need to address due to their fossilization. Therefore, the study was designed around specific segments rather than the overall intelligibility. The participants needed to carry out the spontaneous, controlled, and imitative task to assess their progress after two weeks of daily use of the English File app for 20 minutes. Suzukida and Saito (2022) concentrated on identifying segmental and suprasegmental pronunciation features with high and low functional load in the pronunciation proficiency of Japanese learners of English across different proficiency levels. The participants were evaluated by expert native-speaking raters on a 9-point IELTS rating scale.

As far as social media in pronunciation training are concerned, they appear to be a source of authentic material for pronunciation improvement. In Al-Ahdal's study (2020), the teacher was a mediator and a human corrector of learners' errors in free speeches of students who improved their pronunciation by listening to podcasts. Similarly, Fouz-González (2017) used non-native pronunciation experts to evaluate the pronunciation of commonly mispronounced words after using Twitter.

## **6. Discussion and conclusions**

All selected studies were conducted on adult learners of English; however, their proficiency levels varied.

Concerning the first research question, various specific pronunciation aspects were evaluated. For instance, e.g. Moxon (2021) used automatic speech recognition software Speechace for sounds, Fouz-Gonzalez in 2017 focused on the pronunciation of particular words and, in 2020, on the accuracy of advanced learners' fossilized sounds. Zoghbor (2018) focused on LFC features, and implicitly a similar approach was taken by Al-Ahdal (2020), who corrected learners' pronunciation in discussions on TED Talks. Gluhareva and Prieto (2017) improved speakers' rhythm and comprehensibility.

However, Koet & van den Bergen (2018) used a more holistic approach and analyzed the positive and negative attributes of pronunciation and standardness. Similarly, Suzukida and Saito (2022) evaluated pronunciation features with high and low functional loads for overall comprehensibility.

In terms of actual rating scales and rubrics, the studies tended to be vague, except Suzukida and Saito (2022) used the IELTS Pronunciation Scale and Luo (2014) used a 4-point scale for segmental and suprasegmental levels. The best score was achieved for native-like accuracy, while the lowest was given for incomprehensibility, combining both principles.

Regarding the second research question, the following observations were made. Human and digital tools are used for evaluation – while human raters appear to be efficient in evaluating intelligibility or more complex aspects of pronunciation (e.g., intonation) by various methods (buzzing method, Zoghbar 2017; semantic differential Koet & van den Bergen, 2018), digital tools concentrate on accuracy.

The question of raters' expertise was raised by Koet & van den Bergen (2018), who did not find significant differences between expert and inexperienced listeners; however, they acknowledge that native speakers tend to be more positive in evaluation than in evaluation of non-native speakers. Fouz-Gonzalez (2017 and 2020) used pronunciation experts to evaluate speakers' accuracy, and Gluhareva and Prieto (2017) asked native speakers to evaluate rhythm and comprehensibility. Teachers were excluded from rating by Zoghbor (2018, p. 5) because "English teachers often have exceptionally low thresholds of intelligibility". On the contrary, Luo (2014) used peer evaluation to benefit both learners and peer raters.

As the study presents, the approaches to pronunciation evaluation in pronunciation classrooms are diverse, and peer or expert evaluation can be helpful in the variety.

### *Acknowledgements*

The paper presents partial research outcomes of the projects KEGA 019TTU-4/2021 Introducing new digital tools into teaching and research within transdisciplinary philological study programmes and 7/TU/2021 Pronunciation mistakes of pre-service teachers of English.

## References

- Al-Ahdal, A. (2020). Overcoming pronunciation hurdles in EFL settings: An evaluation of podcasts as a learning tool at Qassim University Saudi Arabia. *Asian EFL Journal Research Articles*, 27.
- Byram, M. & Hu, A. (2017). *Routledge encyclopedia of language teaching and learning*. London and New York, Routledge.
- Celce-Murcia, M., Brinton, D., & Goodwin, J. (1996). *Teaching pronunciation*. Cambridge University, Cambridge.
- Choy, L. T. (2014). The strengths and weaknesses of research methodology: Comparison and complimentary between qualitative and quantitative approaches. *IOSR journal of humanities and social science*, 19(4), 99-104.
- Council of Europe. (2018). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with New Descriptors*. Strasbourg: Council of Europe.
- Fouz-González, J. (2017). Pronunciation instruction through Twitter: the case of commonly mispronounced words. *Computer Assisted Language Learning*, 30(7), 631-663.
- Fouz-González, J. (2020). Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation app. *Language Learning & Technology*, 24(1), 62-85.
- Gluhareva, D., & Prieto, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research*, 21(5), 609-631.
- Hendriks, B., van Meurs, F., & Usmany, N. (2021). The effects of lecturers' non-native accent strength in English on intelligibility and attitudinal evaluations by native and non-native English students. *Language Teaching Research*, 1362168820983145.
- Isaacs, T. & Harding, L. (2017). Research Timeline. Pronunciation assessment. *Language Teaching*, 50, 347-366
- Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, 10(2), 135-159.
- Kelly, G. (2006). *How To Teach Pronunciation (With Cd)*. Pearson Education India.
- Kizlik, B. (2012). Measurement, assessment, and evaluation in education. Retrieved October, 10, 2015.
- Koet, T., & van den Bergh, H. U. U. B. (2018). Pride and prejudice? Judging non-native pronunciation of English. *L1 Educational Studies in Language and Literature*, (Edited by L. Araujo, N. Elf, R. Funke, T. Janssen, J. van Rij, M. Schrijvers, & SK Tse).
- Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual review of applied linguistics*, 27, 184-202.
- Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL quarterly*, 39(3), 369-377.
- Luo, B. (2016). Evaluating a computer-assisted pronunciation training (CAPT) technique for efficient classroom instruction. *Computer assisted language learning*, 29(3), 451-476.
- Moxon, S. (2021). Exploring the Effects of Automated Pronunciation Evaluation on L2 Students in Thailand. *IAFOR Journal of Education*, 9(3), 41-56.
- Munro, M. J. (2010, September). Intelligibility: Buzzword or buzzworthy. In *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching conference* (pp. 7-16).
- Pokrivčáková, S. (2015). 2.2 CALL and teaching pronunciation. *CALL and Foreign Language Education: e-textbook for foreign language teachers*. Pokrivčáková, S. et al. (Eds.). Nitra: Constantine the Philosopher University.
- Scanlan, C. L. (2012). *Assessment, Evaluation, Testing and Grading*. Available online.
- Sheppard, B. E., Elliott, N. C. & Baese-Berk, M. M. (2017). Comprehensibility and intelligibility of international student speech: Comparing perceptions of university EAP instructors and content faculty. *Journal of English for Academic Purposes*. 26, 42-51.
- Suzukida, Y., & Saito, K. (2022). What is second language pronunciation proficiency? An empirical study. *System*, 106, 102754.
- Vančová, H. (2019). Current issues in pronunciation teaching to non-native learners of English. *Journal of Language and Cultural Education*, 7(2), 140-155.
- Vančová, H. (2020). *Pronunciation Practices in EFL Teaching and Learning*. Hradec Králové: University of Hradec Králové, Gaudeamus Publishing House.
- Vančová, H. (2021). Teaching English pronunciation using technology. Nürnberg: Kirsch Verlag.
- Zahra, A., Cabral, J. P., Kane, M., & Carson-Berndsen, J. (2012). Automatic classification of pronunciation errors using decision trees and speech recognition technology. *Proceedings of IS ADEPT, Stockholm*, 65-69.
- Zoghbor, W. S. (2018). Teaching English pronunciation to multi-dialect first language learners: The revival of the Lingua Franca Core (LFC). *System*, 78, 1-14.