

Development of rating scales for the EPPLE examination to assess EFL teachers' oral performance

Desenvolvimento de escalas de proficiência para o exame EPPLE para avaliação de desempenho oral de professores de inglês

Diego Fernando de Oliveira

Universidade Estadual Paulista

Douglas Altamiro Consolo

Universidade Estadual Paulista

Marina Melo Cialdini

Universidade Estadual Paulista

Abstract: The definition of the linguistic aspects and the domain in which to operate to assess language teachers' proficiency, and the criteria to rate teachers' language, constitute challenges in language assessment. In this paper, we describe the main features of the EPPLE (Proficiency Examination for Foreign Language Teachers) oral test, and reflect about results of four research studies: (1) criteria to assess grammatical accuracy and complexity (Martins, 2023; Borges Almeida, 2009a), (2) phonological aspects (Cialdini, 2023), and (3) a proposal of an analytic rating scale for the oral test (Oliveira, 2021) considering five aspects: fluency, grammatical accuracy, metalanguage, pronunciation and vocabulary.

Keywords: Assessment; EFL; EPPLE; Oral proficiency; Rating scales

Resumo: A definição dos aspectos linguísticos e do domínio no qual operacionalizar a avaliação da proficiência linguística de professores de línguas, e dos critérios para avaliar a linguagem de professores, são desafios na área de avaliação. Neste artigo, descrevemos as características do teste oral do EPPLE (Exame de Proficiência para Professores de Línguas Estrangeiras), e refletimos sobre resultados de quatro pesquisas: (1) critérios de avaliação de complexidade e precisão gramatical (Martins, 2023; Borges Almeida, 2009a), (2) aspectos fonológicos (Cialdini, 2023), e (3) faixas de proficiência analíticas para o teste oral (Oliveira, 2021), contemplando fluência, precisão gramatical, metalinguagem, pronúncia e vocabulário.

Palavras-chave: Avaliação; EFL; EPPLE; Escalas de proficiência; Professores de língua inglesa; Proficiência oral

Introduction

In this paper, we reflect about the results of four studies in which linguistic features of spoken language produced by EFL teachers and teachers-to-be were analyzed so as to contribute towards assessment criteria of teachers' oral proficiency: grammatical accuracy and complexity, phonological aspects and a proposal for an analytic rating scale for an oral test. The latter is constituted of five aspects: fluency, grammatical accuracy, metalanguage, pronunciation and vocabulary.

Language proficiency is a requirement for language teachers, and the definition of both the linguistic aspects and the extension of a domain in which to operate to assess this proficiency has been a challenge in language assessment. Teachers' proficiency is understood when linguistic knowledge and the competence for communication lead to effective language use in teaching context for example, in managing classroom language and discourse. Given the relevance of investigating teachers' language, and the connections between the domain, the testing instruments and the criteria on which to base valid assessment for teachers' language proficiency, members of the ENAPLE-CCC,¹ a research group at the UNESP – Sao Paulo State University, in Brazil, have been investigating the aspects of foreign language teachers' proficiency. The studies carried out by the members of this research group have the objective of implementing the EPPL in Brazil (Consolo; Teixeira da Silva, 2007; Consolo *et al.*, 2010), a language examination for foreign language (FL) teachers.

In the next two sections of this paper, we describe the main features of the EPPL oral test, and reflect about research results, in the scope of the examination, namely on criteria to assess grammatical accuracy and complexity in spoken language (Martins, 2023; Borges Almeida, 2009a), a suggested rating scale (Oliveira, 2021) and phonological aspects (Cialdini, 2023). It is expected that the research results presented here can contribute towards more adequate assessment criteria for the EPPL oral test, especially for grammar and pronunciation.

We conclude the article by suggesting implications of these results for the development of more scientifically grounded criteria to assess foreign language teachers' oral proficiency, by means of rating scales.

The EPPL Examination

The EPPL examination stands for Exame de Proficiência para Professores de Línguas Estrangeiras (in Portuguese), a proficiency examination for FL teachers, to evaluate and classify their linguistic proficiency - henceforth LPFLT (language proficiency of foreign language teachers), a type of language proficiency that is both general and specialized (Consolo, 2007).

¹ ENAPLE-CCC stands for Ensino e Aprendizagem de Línguas: crenças, construtos e competências (Language Teaching and Learning: beliefs, constructs and competences), a research group established at UNESP, the Sao Paulo State University, Brazil, and recognized by the CNPq, a Brazilian national sponsoring agency.

The examination aims at teachers-to-be, as well as in-service teachers, that is, undergraduates about to conclude an initial teacher education program at tertiary level, usually in a Letters course (in the case of Brazil), and FL teachers already engaged in the profession and responsible for foreign language lessons to young children, in primary and secondary education, at university and in language schools.

LPFLT includes the abilities of comprehension and production of the foreign language focused on a given version of the EPPL,² in both verbal and written modalities. General language proficiency, regarded as part of the LPFLT, is characterized by the quality of performance in a given language, as it is used by the majority of its speakers in a variety of everyday situations, from informal to formal conversations, when reading informative texts and usual documentation, to understand oral language in verbal messages, and in the production of e-mail messages and written texts aiming at social networking, for example. Considering the specialized proficiency of FL teachers, the main part of LPFLT that mostly determines language proficiency for professional demands, it encompasses the use of a given FL for educational purposes, for example, to manage classroom discourse and communication in language teaching contexts. In this sense, teachers' language includes providing information and giving instructions, pedagogical explanations, evaluating students' performance, reading academic texts and teaching materials, the understanding of audio and videos for pedagogical purposes, and the production of materials and instruments to evaluate students. More detailed reviews and studies about teachers' language have been the focus of members of the EPPL team of researchers and their supervisees, such as Andreilino's PhD dissertation (2014), Colombo's MA thesis (2014) and PhD dissertation (2019), Ducatti's MA thesis (2010), Fernandes' MA thesis (2011) and PhD dissertation (2016).

The examination is constituted of two tests, a paper to assess reading comprehension and written production, and a test of listening comprehension and speaking. The oral test, if administered in its face-to-face format, is conducted with two test takers, in the presence of an examiner who manages the test, and of an examiner who acts more like an observer and a rater. In the computer-based version of the EPPL examination, test takers' do both parts of the examination (the oral test and the written test) in a sequence. Answers are recorded and evaluated by two raters on a later occasion.

In the oral test – in the face-to-face format and in the computer-based format, listening comprehension and oral production skills are integrated. The oral test has four parts. In the first part, test takers are expected to speak about themselves – about personal and professional information, previous experiences as FL students, and professional expectations for the future. The second part of the oral test is based on a brief video segment, or on two short video extracts, that must be understood so that a discussion about the content in the video(s) can be carried out by the test takers with each other (in the face-to-face format), and with the examiner conducting the test. In the computer-based format, the test taker answers comprehension questions and presents her/his

² The EPPL has so far been designed and piloted only in English. However, plans of the ENAPLE-CCC include the production of EPPL in other foreign languages taught in Brazil such as French, German, Italian and Spanish.

comments about the video segment. In the third part of the test the candidates must show their proficiency in using metalanguage, that is specific language for pedagogical purposes. Situations of problems usually faced by FL students are presented and the examinees are expected to offer solutions for linguistic doubts likely to be raised in language lessons. Examinees are expected to explain linguistic rules, for example, as well as produce pedagogical explanations. In the last part of the oral test, test takers are asked for their opinions about the oral test they just did, to provide data about the EPPL from the examinees' perspective.

Electronic versions of the EPPL in English have been produced and administered to student-teachers since 2011. The 'electronic EPPL' is a computer-based examination that includes the tasks for the oral test, to be done in around 25 minutes, and the tasks for the written test, to be done in the second part. The test tasks in the electronic version of the examination are mostly the same, or similar, to its face-to-face version. The whole electronic examination has a maximum duration of one hour and fifty minutes.

The electronic EPPL includes recorded instructions given to the test takers at the beginning of the examination, and an initial task to test the camera and the microphone connected to the computer before the oral test starts. The answers produced by the candidates are recorded in the computers and in a data bank to be rated on a later occasion.

The ratings have so far been based on the holistic rating scale (Appendix 1). Hence, the reflection in this paper, about the research studies aforementioned, aims at helping the definition of analytic rating scales for the oral test. Once the two types of scales are available, each examiner will rate the candidates based on one of the two scales, which shall contribute with more validity and reliability of the EPPL results.

Accuracy, grammatical complexity and pronunciation assessment: The search for valid scales for the EPPL oral test

In this section, we discuss the theoretical bases and research results from four studies that can contribute to the development of criteria to assess foreign language teachers' linguistic oral proficiency, focusing on grammatical complexity and accuracy, a proposed model of an analytic rating scale for the EPPL oral test, and phonological aspects.

In every concept of language assessment and proficiency, concepts of language and grammar are embedded. In line with the definition proposed by Larsen-Freeman (2003), that language is "a dynamic process of pattern formation through which humans use linguistic forms to mean in ways appropriate to the context" (p. 142), this dynamic view proposed by the author highlights that grammar is not a static product governed by rules, a closed and absolute system, but a dynamic process in which forms have meanings and uses in an open, interconnected system. Therefore, in the author's tripartite model, grammar is formed by three dimensions: form, meaning and use. Therefore, "grammaring" is the ability to use grammatical structures accurately, with meaning and appropriately (Larsen-Freeman, 2003).

Luoma (2004) draws attention to the fact that the grammar of orality must be evaluated differently from the grammar of writing, since they are different linguistic manifestations. A very significant first difference between speaking and writing, for example, is the fact that speakers do not speak in sentences, but speak in forms that Luoma (2004) calls idea units, and this occurs, among other reasons, to help listeners understand spoken language in real time. As a result of this temporal limitation, these units tend to be syntactically simpler (with more coordination and less subordination), delimited by intonation contours and marked by hesitations and pauses – characteristic of spoken language.

Grammatical accuracy and complexity in data from TEPOLI and EPPL

Borges Almeida (2009a) conducted a PhD study with graduating students from a foreign language (English) teacher pre-service course (a Letters course) in a state university in Brazil. The study was guided by the research question, “how is grammar characterized across the TEPOLI’s³ proficiency levels in the participants’ performance in terms of (a) accuracy and (b) complexity in the oral test and in a seminar”?

Borges Almeida investigated accuracy in speech production by means of two quantitative indexes: the number of deviations per unit and the percentage of deviation in free-clauses (D’Ely, 2006; Guará-Tavares, 2008). Complexity was investigated in quantitative terms through the mean length of units (given in words), the mean length of clauses, and the frequency of clauses per unit and of dependent clauses per independent unit (Wolfe-Quintero *et al.*, 1998 *apud* Ortega, 2003). The unit of analysis adopted by Borges Almeida (2009a) is Foster *et al.*’s (2000) *AS-unit* (Analysis of Speech Unit), which more clearly presents how to deal with the disfluent mechanisms of speech. An AS-unit is defined as a single speaker’s utterance and it consists of an independent clause, or sub-clausal unit, with any subordinate clause(s) associated with either, and allows for inclusion of independent sub-clausal units, which are common in speech (Foster *et al.*, 2000).

Borges-Almeida (2009a) analyzed oral data from both the TEPOLI and recordings of class seminars presented by the same students who did the test. The indexes indicate that the test has predictive power over a communicative situation that is not an interview in itself – in the case of the seminars, a situation somehow similar to an authentic class. This way, the higher the TEPOLI band, the better the student’s grammatical performance is expected in a non-testing setting as well.

The rating scale for TEPOLI has five bands, from A (the highest) to E (the lowest). The results found by Borges Almeida (2009a) are shown in Table 1 below..

The indexes of self-corrections do not reveal that their occurrences are determined by the proficiency bands, and this raises questions about how self-correction is currently described in the TEPOLI scale, and calls for a revision of this specific item in the descriptors of grammatical competence and performance in spoken language.

³ TEPOLI stands for “*Teste de Proficiência em Língua Inglesa*” (Test of Oral Proficiency in English), a test developed prior to the EPPL examination and piloted with pre-service and in-service EFL teachers in Brazil (Consolo, 2004).

Table 1 – Grammatical accuracy

Band	Index Deviation-Freeunits		Index Deviations per unit		Index Self-correction	
	ORAL TEST	SEMINAR	ORAL TEST	SEMINAR	ORAL TEST	SEMINAR
<i>B</i>	0,88	0,84	0,14	0,19	0,08	0,12
<i>C</i>	0,79	0,67	0,28	0,51	0,09	0,12
<i>D</i>	0,65	0,62	0,46	0,60	0,10	0,06
<i>E</i>	0,61	0,52	0,53	0,97	0,05	0,05

Source: Borges Almeida (2009a)

Grammatical complexity is the “amplitude of forms that emerge from linguistic production and the degree of sophistication of these forms” (Ortega, 2003, p. 492, apud Borges-Almeida, 2009a, p. 87). According to DeKeyser (2005), the complexity of the structures is directly related to the difficulty of acquiring these structures by the learner. The author points out that the relationship between grammatical complexity and grammatical acquisition occurs at three distinct levels, namely the complexity of form, the complexity of meaning and the complexity of the form-meaning relationship.

Kuiken *et al.* (2019) state that speakers of a language can make different lexical, morphological and syntactic choices because of linguistic factors (internal to the language itself) or external. The authors dedicate themselves to investigating the variation in complexity - especially syntactic - due to different external factors. According to these authors, variation can occur due to different levels of proficiency (called developmental variation), modality (oral or written), types of tasks and genres, and also variation in source and target languages. In addition, there are also variations resulting from differences between learners (depending on each one's prior knowledge, interest in the language, personality, etc.) or related to the same learner, such as differences in performance at different stages of learning (Kuiken *et al.*, 2019).

The indexes of complexity (Table 2) presented by Borges Almeida (2009a) suggest that the differences between bands⁴ are small and not much consistent when quantitatively observed. One of the interpretations for the weak consistency of the index of words per unit is that such index can also be employed as a measure of fluency. The index of clauses per unit shows little variation due to its relation to subordination, and subordination is not a characteristic of spoken language. Still the differences found suggest that candidates placed in higher bands tend to produce more complex clauses than candidates placed in lower bands.

Band D can be seen as the level in which the candidates' performance is the most variable. In a paper in which she reports an analysis of data from the TEPOLI from a phonological perspective, Borges-Almeida (2009b) mentions that data from candidates placed in band D for fluency – as analyzed according to the criterium of filled pauses, also show a pattern considerably

⁴ No candidate in the group of TEPOLI takers analyzed by Borges de Almeida (2009a) was placed in band A.

different from those observed in the other bands. One of the hypotheses for that is the one that claims that in this band candidates may go through more linguistic hypotheses restructuring stages, and so band D would be characterized by “a period of variable latency” (Cruz, 2001, p. 51).

Table 2 – Grammatical complexity

Band	Index clauses per unit		Index words per unit	
	ORAL TEST	SEMINAR	ORAL TEST	SEMINAR
B	1,45	1,46	6,62	9,07
C	1,36	1,35	6,79	8,69
D	1,35	0,93	7,32	6,90
E	1,26	1,30	6,32	7,98

Source: Borges Almeida (2009a)

Considering a conception of complexity as structural variety, in this paper complexity is seen from a more targeted and specific perspective, seeking to verify the use of compound verb tenses in the participants' speech (such as the present perfect). In our view, the difficulty of these forms can be found both in the form, given that these tenses need auxiliary verbs, and in the form-meaning relationship that may not be so transparent in some circumstances, as is the case of the perfect aspect, analyzed by Martins (2023).

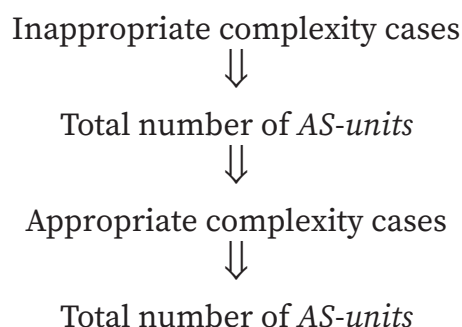
Martins (2023) worked with data from two interrelated but different contexts. For the analysis of grammatical accuracy and complexity, the data in her study were constituted of 29 samples of speech productions from the EPPL oral tests and of 16 video-recorded online English language classes. Each participants' speech was segmented into AS-units (Foster *et al.*, 2000), and the analysis focused on compounded verb tenses. Although AS-units are based on syntax, they consider aspects of oral production such as pauses, false starts, and conversation markers. Norris and Ortega (2009) suggest that because the AS-unit considers non-syntactic elements, it seems to be more appropriate for the analysis of oral data.

In the seminars held during online classes, each student-presenter had to speak about a previously chosen cultural topic, and conduct the presentation as if s/he were teaching an EFL lesson, interacting with other students, asking questions, offering explanations and providing instructions, for example. Therefore, the student-presenter had to make use of metalanguage in his or her speech, and the grammatical complexity in this metalanguage was analyzed.

Grammatical complexity was analyzed from quantitative and qualitative perspectives. Considering a more multidimensional proposal that conceives complexity as structural variety and sophistication, and as a linguistic aspect acquired over time (Norris and Ortega, 2009), Martins (2023) investigated the use of tenses and verbal aspects (more specifically of compounds, namely the past continuous, the present perfect, the present perfect continuous, the past perfect, the past perfect continuous, the future perfect, and the future perfect continuous) in

the speech of the participants in her study, since these are significant for an adequate, complex and precise speech of the FL teacher. In addition, these verb tenses may be more difficult for EFL learners to acquire.

For a quantitative analysis of complexity, Martins (2023) looked for two indices, namely unit-deviation and unit-without-deviation:



According to the unit-without-deviation index, the more complex the student's speech in relation to the chosen criterion, the higher is his/her level of proficiency. On the other hand, the higher the unit-deviation index, the less complex is the student's speech.

The results discussed here showed that the compound tense with which students had most difficulty was the present perfect. In some specific cases, the difficulty is related to form, mainly in the past participle of the verbs. Students had problems in identifying certain moments when it was necessary to use the present perfect, such as when talking about an action or situation that ended in the past without a specific time reference, or to talk about an action or situation that started in the past and extends into the present (usually followed by "for" or "since"). A possible reason for the complexity of this tense is the perfect aspect. DeKeyser (2005) points to the fact that some structures are more difficult to be acquired by speakers whose mother tongue does not have that structure or uses it in a different way. Therefore, as most students seem to be familiar with the past simple, they choose to use this tense instead of the present perfect. In addition, it was possible to verify that, in both contexts, past continuous deviations were rare. And, when necessary, most students used this tense in a precise, meaningful and appropriate way. It is also important to highlight that the use of past perfect continuous, the future continuous, the future perfect and the future perfect continuous was not identified in the EPPL tests or in the online classes.

Based on the results of her study, Martins (2023) elaborated grammatical complexity analytic descriptors for the EPPL's oral test (Chart 1). She produced descriptors for four bands, instead of five bands, because it has been decided, by the members of the ENAPLE-CCC group, that performance in future versions of the EPPL examination will be rated with four-band scales. Four-band scales facilitate the decision to place a test taker either in one of the two lowest bands, which indicates unsatisfactory performance, or in one of the two higher bands, avoiding candidates to be placed in a middle band.

For band 4, the frequency the participant uses the compound verb tenses was considered, as well as accuracy, meaning and appropriateness. For band 3, the accuracy is maintained during at least half of the participant's speaking time. The main difference between bands 2 and 1 is that students classified in band 2 attempted to use complex structures (even if these structures were inaccurate), while there were no such attempts from students classified in band 1.

Chart 1 – Bands and grammar descriptors for the EPPLÉ oral test

Band	Descriptors
4	When necessary, tenses and complex structures are used most part of the time accurately, meaningfully and appropriately.
3	Tends to use complex structures accurately when necessary, at least half the time of his/her speech.
2	May attempt to use more complex structures (such as verb forms) with inadequacies in form, meaning or use.
1	Does not use complex verb tenses, even when prompted by the task.

Source: Martins (2023)

In the next subsection, we report on Oliveira's (2021) study and his suggested model of analytic rating scales for the EPPLÉ oral test.

The proposed analytic scale for the EPPLÉ oral test

The empirically developed analytic scales for the EPPLÉ oral test (Oliveira, 2021) comprises four evaluative categories based on operational definitions by Fulcher (2014) for linguistic competence - pronunciation, grammatical accuracy, fluency, and vocabulary, and a specific category focused on metalinguistic evaluation, since metalinguistic features are an important aspect to be assessed in the EPPLÉ examination. Metalinguistic competence is the use that teachers make of the language to communicate with the students at the same time s/he uses language to talk about the language being taught (Almeida Filho, 1993).

Structured into four levels, the scales' primary aim is to offer descriptive precision to the evaluator concerning the criteria to be considered for the classification of test takers' performances⁵. Although the present holistic scales have five levels, it is desirable to have four levels in both types of scales in order to avoid a middle band. Four-band scales avoid classifying test takers at this "borderline level". As a result, the two lower levels can be seen as inadequate language proficiency for teachers, whereas the two upper levels of the scales represent adequate proficiency levels.

The methodology adopted for scale development involves constructing a decision tree of performances, consisting of a series of binary criteria (yes/no) comprising three main axes. Thus, the presence or absence of certain elements constitutes the criterion delimiting the different levels of the scale. Known as Performance Decision Trees (Fulcher; Davidson; Kemp, 2011), this

⁵ The full analytic scale proposed by Oliveira (2021) is available in Appendix 2.

methodology allows for the consideration of performance data, thereby forming part of the evidence analysis to describe proficiency levels. The database used for the development of the analytic scales results from collections of recordings of oral performances of the EPPL oral test and their transcription, conducted by Colombo (2019).

Regarding the evaluative category of pronunciation, the data demonstrated that the use of intonation in words is an important aspect in defining more and less satisfactory performances. Additionally, at the phonetic level, the pronunciation of segments is also relevant, as its recurrence was related to the number of irregularities in intonation usage, affecting the intelligibility of utterances.

As for grammatical accuracy, it was observed that occurrences of addition or omission of morphosyntactic elements were a recurrent phenomenon and needed to be included in the exam construct. Furthermore, problems with word and sentence formation were also included, as they directly impacted the intelligibility of analyzed utterances.

Concerning fluency, the data showed that the number of pauses increased progressively in relation to the number of other disfluency phenomena, such as restructurings. Thus, the recurrence of pauses constitutes the differentiating criterion for lower levels, and restructurings are the criteria that differentiate more and less satisfactory performances.

Another important aspect observed in the data was the breadth and precision of the vocabulary⁶ used in the analyzed performances. Directly related to the evaluative category of vocabulary, the use of technical terminology in the data constitutes the main criterion for classifying the analyzed data as more or less proficient.

In summary, the analytic rating scales developed by Oliveira (2021) for the EPPL oral test aims at contributing to the improvement of the examination, which prior to the development of this scale lacked more precise descriptors to guide the classification of oral performance. As a result, this study has enabled the expansion of empirical investigations into teacher proficiency, with special attention to the evaluative categories included in the analytic scales.

Criteria to assess phonological aspects in EPPL: improvements in the descriptors of the analytic proficiency bands to assess pronunciation

In this subsection, we present results from Cialdini's MA thesis (2023) and part of the methodology followed in her study.

Cialdini's (2023) study was based on one of Oliveira's (2021) recommendations for future research and the need to analyze the relevance of the criteria found in the analytic scales proposed by Oliveira for the EPPL oral test. Accordingly, Cialdini focused on the criteria that constitute the proficiency bands used to assess pronunciation: 'intonation' to emphasize important elements in the utterance, 'lexical stress' and 'variations' in the production of consonants and and/or vowels. Thus 53 oral samples of (prospective) EFL teachers, from answers in the EPPL oral test and from

⁶ Studies on vocabulary in the assessment of oral performance were conducted by Baffi-Bonvino (2010) and by Consolo and Silva Neto (2015), members of the ENAPLE-CCC research group.

seminar presentations in online lessons – the same data analyzed by Martins (2023), with a few additions –, were assessed according to these criteria. The data in Cialdini's (2023) study is then part of two corpora: 29 recordings of the administration of the EPPLE oral test in 2015 and 2017; and 24 recordings of English lessons in a Portuguese and English undergraduate course (Letters course) at a state university in Brazil, in which students presented seminars in English.

Data were generated using *Phon*, a software tool for analyzing textual and phonological corpora. The query functions available in *Phon* allow the extraction of data that are relevant to the researcher, thus allowing the generation of customized reports.

Some considerations were achieved regarding the criteria for assessing pronunciation in each of the proficiency bands, which can be seen in Table 3 below:

Table 3 – Preliminary considerations about criteria to assess pronunciation in each proficiency band

	Is intonation used?	Consonant and vowel deviations	Word stress deviations	Deviations: consonants and vowels + word stress?
Band 4	Yes	Up to 6	1	No
Band 3	No	Up to 6	Up to 4	No
Band 2	No	5 to 15	1 to 6	Yes
Band 1	No	18 to 58	1 to 12	Yes

Source: Cialdini (2023)

The preliminary considerations presented in Table 3 refer to the quantitative aspects that Cialdini (2023) found to be typical of each proficiency level. We can see that even the participants who were placed in bands⁷ 4 and 3 - i.e., the bands that represent the most proficient performances -, produced deviations in consonants and vowels. Not only did the number of such deviations tend to be more prevalent as proficiency declined, but so did the number of word stress inconsistencies. Occurrences of segmental deviations in simultaneity with word stress deviations were observed only in the oral production samples assigned to bands 2 and 1. On the other hand, segmental deviations produced by participants assigned to band 4 did not affect intelligibility; this finding can be explained not only by the nature of the deviations, but also by the adequate control of suprasegmental features (fluency and rhythm) demonstrated by these participants, which made such inaccuracies virtually imperceptible. For example, the participant JOV⁸ pronounced the word 'chose' as ['ʃoz] instead of ['ʒoz].

Furthermore, it was found that the lower the proficiency band, the greater the impairment of segmental deviations for intelligibility. For instance, the word 'scissors' was pronounced by the

⁷ Cialdini (2023) also established assessment criteria in a four-band scale.

⁸ Identification acronyms refer to participants' first names and either their middle names or surnames.

participant ISV (band 2) as ['saizəz] instead of ['sizəz], while DEI (band 1) pronounced the word 'know' as ['naʊ] instead of ['nou]. L1 (Brazilian Portuguese) transfer was also observed as a typical feature of the participants whose oral performance was assigned to band 1, like the epenthetic [ɪ] in consonantal ending words, e.g. DEI's pronunciation of 'England' as ['ɪŋɡlɪndɪ] instead of ['ɪŋɡlɪnd].

In order to obtain additional information that would confirm and complement these preliminary considerations, four assessment specialists were invited to contribute to Cialdini's (2023) study. Their contribution consisted in evaluating the oral production samples belonging to the second corpus (the seminars presented by 29 undergraduate students in a Letters Course). These raters were instructed to use the analytic proficiency bands elaborated by Oliveira (2021) to assess pronunciation, and also to leave comments that justified their ratings. The total number of samples was distributed among these raters so that each sample ended up being assessed by three raters, one of whom was Cialdini (2023). The ratings assigned by these experts to the aforementioned samples were compared with the preliminary ratings assigned by Cialdini (2023).

Disagreements between the raters can be attributed to the vagueness of the adjective 'frequent' (bands 2 and 1) and the criterium 'intonation to emphasize important elements in the utterance' (bands 4 and 3). Difficulties with the interpretation of this criterium were expressed by three of the raters.

Disagreements between the four assessment specialists and Cialdini (2023) can be attributed to the fact that Cialdini's analysis was based on both quantitative and qualitative aspects, whereas the raters' assessment was essentially qualitative. However, discrepancies in the ratings between the scope of bands 4 and 3, and between the scope of bands 2 and 1 were not treated as concerning, as they belong to the same categories (bands 4 and 3 correspond to 'satisfactory' performances, while bands 2 and 1 refer to performances that are considered 'unsatisfactory' by EPPLÉ specialists).

Cialdini's (2023) research has resulted in a revision of the analytic proficiency bands used to assess pronunciation in EPPLÉ, as follows.

According to Chart 2, the criteria used for the assessment of pronunciation within the proficiency bands, as redefined by Cialdini (2023), are presented in such a manner that considers a parallel relation between the descriptors under the same numbers (1-6). Also, descriptors are presented progressively, and new phonological aspects have been added as criteria.

The results from the four studies aforementioned (Cialdini, 2023; Martins, 2023; Oliveira, 2021; Borges Almeida, 2009a) can help the inclusion of more objective and precise, and scientifically based descriptors as criteria for language assessment in future analytic rating scales for the EPPLÉ oral test. And it is desirable that the holistic scale is revised and rewritten in line with the analytic scales.

Chart 2 – Revised analytic proficiency bands to assess pronunciation

4	<ol style="list-style-type: none"> 1. Speech is intelligible and comprehensible throughout. 2. Uses intonation to highlight important elements in his / her utterances. 3. Deviations in the production of consonants and vowels are rare. 4. Deviations in word stress are virtually non-existent. 5. Deviations in the production of consonants and vowels are not observed concomitantly with deviations in word stress. 6. Has satisfactory control over the articulation of complex sounds of the target language.
3	<ol style="list-style-type: none"> 1. Speech is intelligible and comprehensible throughout. 2. Does not use intonation to highlight important elements in his / her utterances. 3. Deviations in the production of consonants and vowels are rare. 4. Deviations in word stress are rare. 5. Deviations in the production of consonants and vowels are not observed concomitantly with deviations in word stress. 6. Has moderate control over the articulation of complex sounds of the target language.
2	<ol style="list-style-type: none"> 1. Speech is occasionally unintelligible and incomprehensible. 2. Does not use intonation to highlight important elements in his / her utterances. 3. Deviations in the production of consonants and vowels are generally observed. 4. Deviations in word stress are occasionally observed. 5. Deviations in the production of consonants and vowels are observed concomitantly with deviations in word stress. 6. Has reasonable control over the articulation of complex sounds of the target language.
1	<ol style="list-style-type: none"> 1. Speech is mostly unintelligible and incomprehensible. 2. Does not use intonation to highlight important elements in his / her utterances. 3. Deviations in the production of consonants and vowels are frequent. 4. Deviations in word stress are frequent. 5. Deviations in the production of consonants and vowels are observed concomitantly with deviations in word stress. 6. Has little control over the articulation of complex sounds of the target language.

Source: Cialdini (2023)

Conclusion

In this paper, we presented a brief description of the EPPL examination, and results from investigations in which this examination provided data to support the design of more precise rating scales, thus helping the implementation of the EPPL examination and of its construct.

The process of designing the EPPL and establishing its implementation in the context of educating FL teachers in Brazil is on the way towards a revised construct for the examination, the development of more valid test items and the definition of assessment criteria informed and supported by research studies such as the ones presented here. As a follow-up to the studies aforementioned, discussions are to be conducted so as to align the existing holistic scale, the scale proposed by Oliveira (2021) and other research results, and develop a revised version of the holistic scale and corresponding analytic scales for the EPPL oral test. The combined use of the holistic and analytic scales will contribute to more consistent oral performance ratings, thus increasing the validity and reliability of the results of the examination.

Once the EPPL examination is largely available, it is expected to motivate new definitions of course contents and aims in pre-service and in-service teacher education, and may be considered a reference for the quality of language teaching and learning in the Brazilian educational contexts.

References

- ALMEIDA FILHO, José Carlos Paes de. **Dimensões Comunicativas no Ensino de Línguas**. Campinas, Brazil: Pontes Editores, 1993.
- ANDRELINO, Paulo José. **Análise da Estrutura Genérica das Instruções na Fala do Professor de Inglês: contribuições para o teste oral do EPPL** (An Analysis of the Generic Structure in English Teachers' Talk: Contributions to the EPPL oral test). Unpublished PhD dissertation. Sao Jose do Rio Preto, Brazil: UNESP, 2014.
- BAFFI-BONVINO, Melissa Alves. **Avaliação da proficiência oral em inglês como língua estrangeira de formandos em Letras: uma proposta para validar o descritor 'vocabulário' de um teste de professores de língua inglesa** (The Assessment of Oral Proficiency in English as a Foreign Language of Graduating Students in Letters: A proposal to validate the descriptor 'vocabulary' in a test for English language teachers). Unpublished PhD dissertation. São Jose do Rio Preto, Brazil: UNESP, 2010.
- BORGES ALMEIDA, Vanessa. **Precisão e Complexidade Gramatical na Avaliação de Proficiência Oral em Inglês do Formando em Letras: implicações para a validação de um teste** (Grammatical Precision and Complexity in the Assessment of Oral Proficiency of Letters Graduating Students: Implications for a test validity). Unpublished PhD dissertation. São Jose do Rio Preto: UNESP, 2009a.
- BORGES ALMEIDA, Vanessa. Pausas preenchidas e domínios prosódicos: evidências para a validação do descritor fluência em um teste de proficiência oral em língua estrangeira (Filled pauses and prosodic domains: Evidence for the validation of the descriptor for fluency in a foreign language oral proficiency test). **ALFA: Revista de Linguística**, v. 53, n. 1, p. 167-193, 2009b.

CIALDINI, Marina Melo. **Uma Proposta de Aprimoramentos para Faixas de Proficiência destinadas à Avaliação da Pronúncia em um Exame para Professores de Línguas Estrangeiras: um produto da análise dos seus critérios avaliativos e da pronúncia dos sons /θ/, /ð/ e [t] em falas de (futuros) professores de língua inglesa** (A proposal of improvements for analytic proficiency bands to assess pronunciation in an examination for foreign language teachers: A product of the analysis of its criteria and of the pronunciation of the sounds /θ/, /ð/ e [t] in oral production samples of (future) EFL teachers). Unpublished MA thesis: São Jose do Rio Preto, Brazil: UNESP, 2023.

COLOMBO, Camila Sthéfanie. **Avaliação de Proficiência Oral em Língua Estrangeira: Uma Proposta de Abordagem Avaliativa em um Exame para Professores de Línguas com base em Métodos de Estruturação de Problemas e em Métodos Multicritério de Análise da Decisão** (Oral Assessment in a Foreign Language: A proposal of an assessment approach in an examination for language teachers based in methods of problem structuring and methods multicriteria of the analysis of decision). PhD dissertation. Sao Jose do Rio Preto, Brazil: UNESP, 2019.

COLOMBO, Camila Sthéfanie. **O Insumo Linguístico Oral em Aulas de Inglês como Língua Estrangeira para Crianças: a fala do professor em foco** (The Oral Linguistic Input in EFL Lessons for Children: Focus on teacher talk). Unpublished MA thesis. Sao Jose do Rio Preto, Brazil: UNESP, 2014.

CONSOLO, Douglas Altamiro. A competência oral de professores de língua estrangeira: a relação teoria-prática no contexto brasileiro (The oral competence of foreign language teachers: The theory-practice relationship). In: CONSOLO, Douglas Altamiro; TEIXEIRA DA SILVA, Vera Lucia. (Eds.) **Olhares sobre Competências do Professor de Língua Estrangeira: da Formação ao Desempenho Profissional**, Sao Jose do Rio Preto, Brazil: Editora HN, 2007, p. 165-178.

CONSOLO, Douglas Altamiro. (2004). A construção de um instrumento de avaliação da proficiência oral do professor de língua estrangeira. **Trabalhos em Linguística Aplicada**, v. 43, n. 2, p. 265-286, 2004.

CONSOLO, Douglas Altamiro; LANZONI, Hécio de Pádua; ALVARENGA, Magali Berçante; CONCÁRIO, Marcelo; MARTINS, Teresa Helena Buscato; TEIXEIRA DA SILVA, Vera Lúcia. Exame de Proficiência para Professores de Língua Estrangeira (EPPL): proposta inicial e implicações para o contexto brasileiro (Proficiency Examination for Foreign Language Teachers: Initial proposal and implications for the Brazilian context). Proceedings of the **II Congresso Latino-Americano de Formação de Professores de Línguas (CLAFPL)**, Rio de Janeiro, Brazil: PUC-Rio (CD-ROM), 2010.

CONSOLO, Douglas Altamiro; SILVA NETO, Teucle Maurílio. An analysis of lexicon and oral language proficiency in English of Letters undergraduates. **Revista de Letras Norte@mentos**, v. 8, p. 53-78, 2015.

CONSOLO, Douglas Altamiro; TEIXEIRA DA SILVA, Vera Lúcia. The TEPOLI test: Construct, updated tasks and new parameters to assess EFL teachers' oral proficiency. Proceedings of the **I Congresso Internacional da ABRAPUI** (ABRAPUI International Conference). Belo Horizonte, Brazil: UFMG, (CD-ROM), 2007.

CRUZ, Maria de Lurdes Otero Brabo. **Estágios de interlíngua: estudo longitudinal centrado na oralidade de sujeitos brasileiros aprendizes de espanhol** [Stages of interlanguage: A longitudinal study focusing the oral production of Brazilian learners of Spanish]. Unpublished PhD dissertation. Campinas, Brazil: UNICAMP, 2001.

D'ELY, Raquel Carolina Souza Ferraz. **A focus on learners' metacognitive process: the impact of strategic planning, repetition, strategic planning plus repetition, and strategic planning for repetition on L2 performance**. Unpublished PhD dissertation. Florianopolis, Brazil: UFSC, 2006.

DeKEYSER, Robert. What makes learning second-language grammar difficult? A review of issues. **Language Learning**, v. 55, n. 1, p. 1-25, 2005.

DUCATTI, Ana Lúcia Fonseca. **A Interação Verbal e a Proficiência Oral na Língua-Alvo na Prática de Sala de Aula: (re)definindo o perfil do professor de uma professora de língua inglesa da escola pública** (Verbal Interaction and Oral Proficiency in the Target Language in the Classroom Context: (Re)defining the professional profile of an English teacher in a public school). Unpublished MA thesis. Sao Jose do Rio Preto, Brazil: UNESP, 2010.

FERNANDES, Aline Mara. **A (meta)linguagem para explicação gramatical em língua inglesa: subsídios para a elaboração de tarefas do teste oral do EPPE (Exame de Proficiência para Professores de Língua Estrangeira)** (Metalanguage to explain grammar in English: Subsidies to design tasks of the EPPE (Proficiency Examination for Foreign Language Teachers)). Unpublished PhD dissertation. Sao Jose do Rio Preto, Brazil: UNESP, 2016.

FERNANDES, Aline Mara. **A Metalinguagem e a Precisão Gramatical na Proficiência Oral de Duas Professoras de Inglês como Língua Estrangeira** (Metalanguage and Grammatical Accuracy in the Oral Proficiency of Two English as a Foreign Language Teachers). Unpublished MA thesis. Sao Jose do Rio Preto, Brazil: UNESP, 2011.

FOSTER, Pauline; TONKYN, Alan; WIGGLESWORTH, Gillian. Measuring spoken language: a unit for all reasons. **Applied Linguistics**, v. 21, n. 3, p. 354-375, 2000.

FULCHER, Glenn. **Testing second language speaking**. New York: Routledge, 2014.

FULCHER, Glenn; DAVIDSON, Fred; KEMP, Jenny. Effective rating scale development for speaking tests: Performance decision trees. **Language Testing**, v. 1, n. 28, p. 5-29, 2011.

GUARÁ-TAVARES, Maria da Glória. **Pre-task planning, working memory capacity, and L2 speech performance**. Unpublished PhD dissertation. Florianopolis, Brazil: UFSC, 2008.

KUIKEN, Folkert; VEDDER, Ineke; HOUSEN, Alex; CLERCQ, Bastien de. Variation in syntactic complexity: Introduction. **International Journal of Applied Linguistics**, v. 29, p. 161-170, 2019.

LARSEN-FREEMAN, Diane. **Teaching Language: from grammar to grammaring**. Boston: New House Teacher Development, 2003.

LUOMA, Sari. The nature of speaking. In: LUOMA, Sari. **Assessing Speaking**. Cambridge: Cambridge University Press, p. 9-28, 2004.

MARTINS, Ana Cláudia. **A precisão e a complexidade gramatical em língua inglesa: contribuições à avaliação da proficiência linguística oral do professor de língua estrangeira** (Accuracy and grammatical complexity in English: contributions to the assessment of foreign language teachers' oral linguistic proficiency). Unpublished MA thesis. Sao Jose do Rio Preto, Brazil: UNESP, 2023.

NORRIS, John; ORTEGA, Lourdes. Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. **Applied Linguistics**, n. 30, p. 555-578, 2009.

OLIVEIRA, Diego Fernando de. **Faixas de Proficiência Analíticas para Avaliação da Língua(gem) Oral do Professor de Língua Estrangeira** (Analytic Rating Scales to Assess Foreign Language Teachers' Spoken Language). MA thesis. Sao Jose do Rio Preto, Brazil: UNESP, 2021.

ORTEGA, Lourdes Syntactic complexity measures and their relationship to L2 proficiency: a research synthesis of college-level L2 writing. **Applied Linguistics**, v. 24, n. 4, p. 492-518, 2003.

THOMSON, Ron. *Measurement of accentedness, intelligibility, and comprehensibility*. In: KANG, Okim; GINTHER, April (Eds.). **Assessment in Second Language Pronunciation**. New York: Routledge, 2018, p. 11-29.

WOLFE-QUINTERO, Kate; INAGAKI, Shunji; KIM, Hae-Young. Second language development in writing: measures of fluency, accuracy and complexity. Technical report 17. Manoa, Hawaii, US: University of Hawaii Press, 1998.