

LANGUAGE ASSESSMENT IN THE NORMALIZING SOCIETY

Diego Fernando de Oliveira¹

diego.fernando@unesp.br

SÃO PAULO STATE UNIVERSITY (BRAZIL)

ABSTRACT. The aim of this study is to discuss linguistic norms present in the *Common European Framework*, particularly its criteria for assessing phonetic and phonological features in oral performances. For this purpose, we employed Performance Decision Trees to demonstrate that its operational definition for oral proficiency punishes the presence of foreignness in linguistic performances, even when segmental features of the speech do not affect candidates' performances in linguistic tasks. As a result, we observed that the *Common European Framework* progressively penalizes the presence of phonetic and phonological elements associated with other languages spoken by candidates.

KEYWORDS. Language Assessment, Common European Framework, Pronunciation, Phonological Control.

RESUMO. Este estudo tem como objetivo discutir as normas linguísticas presentes no *Quadro Comum Europeu*, atribuindo especial atenção aos seus critérios para a avaliação de aspectos fonéticos e fonológicos em desempenhos orais. Para isso, empregamos Árvores de Decisão de Desempenho a fim de demonstrar que a sua definição operacional de proficiência prejudica a presença da estrangeiridade nos desempenhos orais, mesmo quando esta não afeta o desempenho proficiente dos sujeitos em tarefas linguísticas. Como resultado, observamos que o *Quadro Comum Europeu* pune progressivamente a presença de elementos de natureza fonética e fonológica associados a outras línguas faladas por candidatos.

PALAVRAS-CHAVE. Avaliação de Línguas, Quadro Comum Europeu, Pronúncia, Controle Fonológico.

¹ Second-year doctoral student in the Postgraduate Program in Language Studies at São Paulo State University (Unesp), Brazil.

1. Introduction

We are entering the age of the infinite examination and of compulsory objectification.

(Foucault 1995: 189)

Applied Linguistics' (AL) literature traditionally defines language assessment as a technical process that aims to collect data and measure evidence about a subject's ability to perform language tasks that simulate real-life contexts (Segat & Sarmiento 2022). Another common aspect in specialized literature is the long-standing taxonomy that classifies assessment as formative and summative. Formative assessment focuses on local contexts and pedagogical purposes, while summative assessment has as its main objective the certification of skills and knowledge at the end of a given school term. There is also standardized assessment, which specialists define as a process that seeks to gauge the level of proficiency of candidates, usually associated with international mobility settings (Fulcher 2010).

Despite its conservative perspective, specialized literature has raised debates about the social scope of assessment practices, paying special attention to issues related to the fairness and validity of assessment processes, as well as their impact on promoting social justice (Fulcher & Davidson 2007; Fulcher 2010, 2015; Fulcher & Harding 2021). Thus, assessment studies raise debates that challenge the long-standing taxonomies that confine assessment processes to institutional practices, demonstrating the social and political scope of contemporary assessment practices.

In this regard, McNamara (2011) and Harding & Mcnamara (2018) acknowledge the fact that language assessment praxis in the 21st century demands reflection. According to McNamara (2011: 500),

Language assessment, as we are increasingly realizing, is first and foremost a social and political activity [...]. It is a site where social values are expressed and contested. As the values of a globalized economy have come to occupy and preoccupy many spheres of social and political life, so have they become central to current developments and challenges in language testing and assessment.

Regardless of raising important questions about language assessment processes, literature is often limited in its sociological perspective related to the social impacts of assessment. Assessing languages is also assessing identities and cultures, since these are inseparable elements (Darvin & Norton 2015, 2017). In a world of increasingly liquid boundaries but progressively more controlled spaces, linguists must comprehend language

assessment as a social dynamic. In this respect, this paper aims to demonstrate that institutions can potentially employ language assessment as an instrument of normalization: not only normalization of linguistic performances, but also of linguistic identities.

In view of the context in which this study is set, as well as the precedents in AL that are concerned with discussing the social impacts of language assessment, this study aims to discuss linguistic norms present in the *Common European Framework* (CEFR) companion volume (Council of Europe 2020), specifically to its phonetic and phonological criteria. For this purpose, we employed Performance Decision Trees (PDTs) to demonstrate that its operational definition of language proficiency punishes the presence of foreignness in oral performances, even when it does not affect candidates' proficient performances in linguistic tasks.

As far as the structure of this study is concerned, we have structured our discussions into five main sections: the first section contextualizes our object of analysis, the assessment criteria for oral performances in the CEFR. The second section briefly discusses the intersection between phonetics, phonology and assessment. The third section introduces Foucault's sociological ideas about normalization. The fourth section deals with the research methodology adopted, the PDTs. Finally, the fifth section presents the results, discussing them in the light of the discussions raised in the theoretical sections.

2. The CEFR as a standard for language assessment

The CEFR (Council of Europe 2020) is one of the most successful policies of the Council of Europe, setting standards for language assessment in different parts of the world. According to the Council of Europe, the CEFR

[...] is intended to promote quality plurilingual education, facilitate greater social mobility and stimulate reflection and exchange between language professionals for curriculum development and in teacher education. Furthermore the CEFR provides a metalanguage for discussing the complexity of language proficiency for all citizens in a multilingual and intercultural Europe, and for education policy makers to reflect on learning objectives and outcomes that should be coherent and transparent (Council of Europe 2020: 11)

As can be seen, the Council of Europe emphasizes the role of the CEFR as a language policy open to plurilingualism, multilingualism, and interculturalism. There are several passages in the referred document that portray the CEFR as an instrument that reflects the

plurality of languages, “[...] including the promotion of reflective learning and learner autonomy” (Council of Europe 2020: 13). As such, the CEFR is presented in this document as an assessment tool that is open to diversity, a political statement that meets the demands of decolonial studies in AL.

Taking into account this concern with progressive debates in the field of foreign language teaching and learning, there is also an attempt to construct a transparent image of the linguistic and educational policies described in the document. According to the Council of Europe (2020: 27)

The CEFR was developed as a continuation of the Council of Europe’s work in language education during the 1970s and 1980s. The CEFR “action-oriented approach” builds on and goes beyond the communicative approach proposed in the mid-1970s in the publication “The Threshold Level”, the first functional/notional specification of language needs. The CEFR and the related European Language Portfolio (ELP) that accompanied it were recommended by an intergovernmental symposium held in Switzerland in 1991. As its subtitle suggests, the CEFR is concerned principally with learning and teaching. It aims to facilitate transparency and coherence between the curriculum, teaching and assessment within an institution and transparency and coherence between institutions, educational sectors, regions and countries.

Furthermore, the Council of Europe (2020: 28) states that its main aim is “[...] facilitating quality in language education and promoting a Europe of open-minded plurilingual citizens” and it “[...] brings a new, empowering vision of the learner”. Although statements of this nature proliferate throughout the companion volume, several authors present questions regarding different aspects of the CEFR. The absence of clarity in their descriptors, the presence of culture-specific bias and lack of attention to contexts and purposes of language use and assessment are examples of demands addressed by the CEFR companion volume (Council of Europe 2020).

In addition, various studies highlight challenges in the application of the CEFR, particularly in social contexts that diverge significantly from the European linguistic milieu (Çagatay & Gurocak 2016; Mohammed, Raof & Yusof 2021; Zaki & Darmi 2021, among others). We attribute this to the fact that the CEFR not only serves as an operational framework for the language construct but also embodies the language policy of European social institutions (Mcnamara 2011), which causes dissonance with the realities of South American, African and Asian societies, for example.

Regarding the implications of using the language policies of the CEFR, McNamara (2011: 506-507) states that

The imposition of a single set of cultural meanings and social and political values for language education, for each setting in which the CEFR is adopted, eviscerates the traditions of language teaching which are incompatible with the CEFR. In cultural and historical terms, learning English is simply not the same for a Singaporean, an Indonesian, a Vietnamese, a French person, a Dutch person, or a Hungarian. And different languages are indeed different, and vary in socio-political role, range and purpose of use, and carry very different histories of contact, often violent, between speakers of the speech communities involved. The CEFR puts all foreign languages into one and the same category, thereby erasing the fact that they are ‘foreign’ in very different ways. Learners from those backgrounds may be aware of the cultural and political significance of the act of learning the language of a cultural group with a particular historical relationship to their own, unless they are invited, in an act of collective historical amnesia, to wipe the slate clean and re-identify themselves as citizens of the new globalized world.

As can be seen, despite the position of the CEFR as an instrument developed from a perspective that approaches the diversity of languages in its theoretical scope, its implementation entails a number of problems, especially in societies that are culturally distant from Europe. Language assessment policies such as the CEFR affect different dimensions of people’s lives, which has consequences for access to education, work and social mobility itself (Spolsky 1995; Davies 2008; Mcnamara 2011).

Last, but certainly not least, in introducing the scale, the Council of Europe (2020) points to improvements in the clarity of criteria compared to the 2001 original version of their framework. To this end, regarding pronunciation assessment criteria, the Council of Europe (2020) states that it consulted AL literature and language experts to operationalize language proficiency with greater scientific validation, contemplating the following empirically verifiable aspects of its framework:

Articulation, including pronunciation of sounds/phonemes; prosody, including intonation, rhythm and stress – both word stress and sentence stress – and speech rate/chunking; accentedness, accent and deviation from a “norm”; intelligibility, accessibility of meaning for interlocutors, covering also the interlocutors’ perceived difficulty in understanding (normally referred to as “comprehensibility” (Council of Europe 2020: 133).

In this sense, there is a concern to find scientific evidence to support the standardizing role of the CEFR's language policies. The Council of Europe (2020) further emphasizes the way in which it conceives the elements operationalized on its framework:

The focus is on familiarity and confidence with the target language sounds (the range of sounds a speaker can articulate and with what degree of precision). The key concept operationalised in the scale is the degree of clarity and precision in the articulation of sounds. The focus is on the ability to effectively use prosodic features to convey meaning in an increasingly precise manner (Council of Europe 2020: 133).

What becomes evident is the presence (almost imperceptible) of a norm that supports all the criteria mentioned by the CEFR. In addition, the norm for phonological control refers precisely to the assessors' socially constructed perspective towards the construct of language proficiency. Concerning the assessor's influence over the assessment process, the literature points out that rater training would be necessary to provide more reliability in the results of the assessment process (Fulcher 2010), which does not mean that this strategy is truly effective. However, behind the scientificism of language assessment dynamics, there is the noise of human subjectivity that assessment institutions and instruments cannot erase.

Considering the political, educational and linguistic aspects described by the Council of Europe (2020) in relation to the CEFR, the next section explores the criteria for assessing phonological aspects in the AL literature.

3. Phonology and language assessment and the problem of the nativeness

In order to assessment take place, assessment institutions operationalize language according to models developed by specialists (Fulcher 2010). Such models of language proficiency are constantly being expanded in order to translate proficiency accurately into observable elements in oral performance. Once materialized in criteria, these observable aspects compose descriptors in proficiency scales. In this manner, assessment contexts and purposes determine the operationalization of proficiency, since the results usually guide assessors and institutions to make decisions based on the candidates' performances.

Phonological control is an assessment category that composes a broader field, more commonly referred to in the literature as pronunciation (Fulcher 2010). In order to explain what we are referring to, we have adopted Fulcher's definition of pronunciation, according to which “[...] the outer manifestation of speech is sound. The speaker must first decide what to say, be able to articulate the words, and create the physical sounds that carry meaning” (Fulcher 2010: 25). However, regarding pronunciation assessment practices, Fulcher (2010: 25) states that:

When we design speaking tests we must decide whether assessing pronunciation at this level is relevant at all to the situation. This depends upon the test purpose. Pronouncing words in the way they would be pronounced in the standard variety of a language may be important to a newsreader, but for most learners testing pronunciation may only be a matter of general intelligibility.

As a result, in cases where pronunciation is a relevant language feature for decision-making, the assessment of oral performances takes into account segmental and suprasegmental dimensions of speech (Fulcher 2010). Segmental elements of oral performance refer to vowels, consonants, and their respective singularities. The suprasegmental elements refer to word stress, intonation, tone, rhythm, among other aspects (O'Brien 2021). Such operationalization can be observable in the CEFR criteria for assessing phonological control - and, in this paper, we focus on its hierarchical organization.

Another relevant aspect for assessing pronunciation is intelligibility. According to Kennedy & Trofimovich (2008) and Thomson (2018), intelligibility is the extent to which the listener understands a sentence. Julkowska & Cebrian (2015) conceive intelligibility as the degree to which a speaker's output is actually understood by the listener. Thomson (2018) also points out that in recent works related to language assessment, authors tend to measure intelligibility from segmental aspects rather than at the utterance level. According to Bundgaard-Nielsen, Best, Kroos & Tyler (2012) and Gooch, Saito & Lyster (2016), a segment within a word is intelligible if it is identified by the listener as belonging to the target language system.

The theoretical criteria which references segmental and suprasegmental aspects of speech recognizable to the interlocutor as a part of the "phonological system" of a given foreign language evidence a formalist perspective of language, still considerably dominant in different fields of study. Such a perspective is based on the linguistic fallacy of the existence of a language as a homogeneous system spoken by an idealized native speaker (Camargo 2016), which includes phonological features in oral performances.

In this respect, according to Pennycook (2017), the perspective of a language as a people's heritage owned by native speakers relates to the formation of European states in the Renaissance period. Language, in this sense, is homogeneous, invariable, governed by rules to be followed and not questioned, which must be learned in order to unify a national identity. In this context, Machado, Delfino & Rodrigues (2021: 117, our translation) argue that

conceiving language as a plastic, malleable structure, dependent on actual speakers and real speakers and socio-cultural interactions, which presents regional, social, age, gender, etc. linguistic variations and gender, even though there are many lexical, phonetic and grammatical elements that are stable and common to the varieties of a language, implies questioning colonialist epistemology, valuing plurality and different ways of being, know and produce knowledge of the subjects.

In the case of the CEFR criteria for phonological control, assessing pronunciation in terms of intelligibility puts the performance of the assessed at the mercy of an idealized social norm, often not well described or defined. Assessors, based on their perspective of language and pronunciation, classify the assessed performances as satisfactory or unsatisfactory: if satisfactory, the assessor crowns the norm that they define as the standard variety of the foreign language assessed; if unsatisfactory, the assessor penalizes candidates based on particular views related to ideal speakers.

In view of the discussions raised in this section about the assessment of phonological aspects and the problem of nativeness, the next section explores concepts from sociology that are important for understanding our perspective on assessment as a normalizing dynamic.

4. Norms, normalization and the examination

In this section, we explore sociological concepts that are important for understanding the discussions we are proposing in relation to our conception of assessment as a normalizing process. Firstly, it is important to establish notions of norms and normalization, which guide the objective of this study. Despite not providing a single definition for his concept of norms, we base ourselves on the ideas developed by Foucault (1987) interpretations of the norm as an ideal social model, taking as a reference the study by Kelly (2019), a specialist in Foucauldian studies.

In this way, we consider norms and normalization to be “[...] a model of perfection that operates as a guide to action in any particular sphere of human activity, and normalization correlatively as the movement by which people are brought under these norms” (Kelly 2019: 2). Based on this, the construct of language proficiency operationalized by policies and assessment instruments represents the norm and normalization, in turn, deals with the process in which candidates subject themselves to language assessment criteria.

In addition, Foucault states that normalization consists

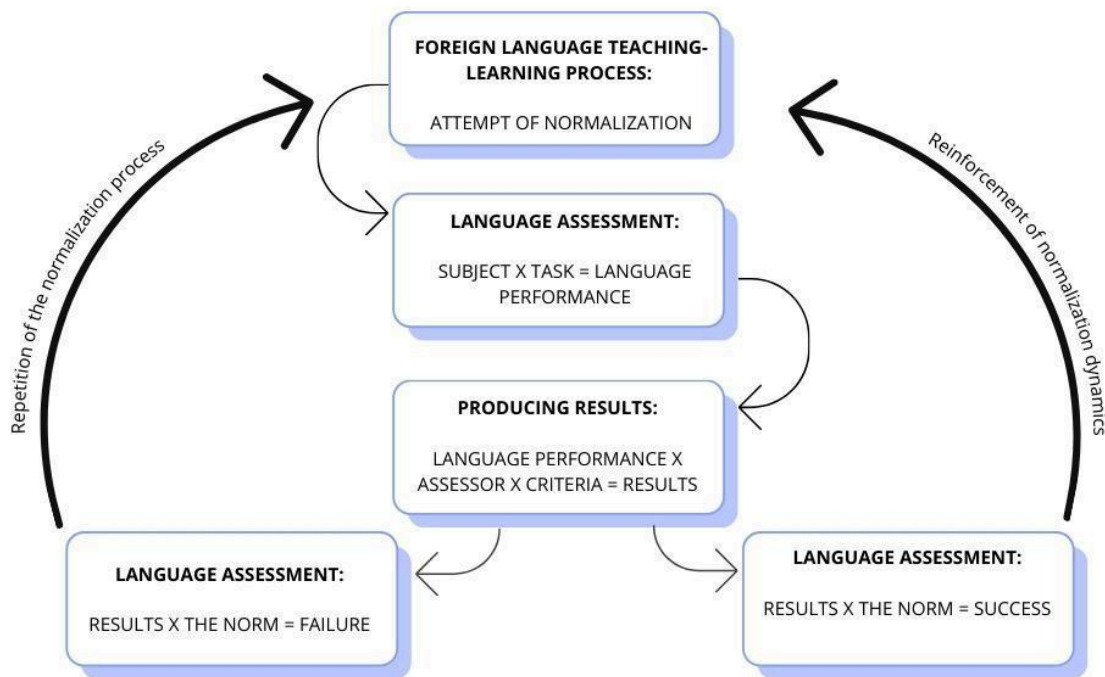
first of all in positing a model, an optimal model that is constructed in terms of a certain result, and the operation of disciplinary normalization consists in trying to get people, movements, and actions to conform to this model, the normal being precisely that which can conform to this norm, and the abnormal that which is incapable of conforming to the norm. In other words, it is not the normal and the abnormal that is fundamental and primary in disciplinary normalization, it is the norm. That is, there is an originally prescriptive character of the norm and the determination and the identification of the normal and the abnormal becomes possible in relation to this posited norm (Foucault 2007: 58).

In his writings, Foucault (1987) further discusses the normalization of subjects through school pedagogical action, referring to what he calls the examination. According to Foucault,

The examination combines the techniques of an observing hierarchy and those of a normalizing judgment. It is a normalizing gaze, a surveillance that makes it possible to qualify, to classify and to punish. It establishes over individuals a visibility through which one differentiates them and judges them. That is why, in all the mechanisms of discipline, the examination is highly ritualized. In it are combined the ceremony of power and the form of the experiment, the deployment of force and the establishment of truth. At the heart of the procedures of discipline, it manifests the subjection of those who are perceived as objects and the objectification of those who are subjected. The superimposition of the power relations and knowledge relations assumes in the examination all its visible brilliance (Foucault 1987: 184).

Based on Foucault (1987, 2007) and Kelly (2019), we interpret the CEFR's linguistic and assessment policies as a series of norms aimed at describing the ideals of linguistic performance to be achieved by candidates who undergo assessment processes. Assessment processes, in turn, categorize, classify and hierarchize subjects according to criteria that punish the presence of linguistic elements considered abnormal within certain idealized standards of foreign languages. We therefore illustrate our conception of assessment as a normalization process in Figure 1 below.

Figure 1 - Language assessment process.



Source: Elaborated by the author.

As can be observed, exam assessment processes based on frameworks such as the CEFR criteria propose linguistic tasks to their candidates in order to collect evidence of their linguistic proficiency — or what we call their linguistic normalization. As a result, subjects are classified as more or less proficient hierarchically: those who succeed in this process have access to socioeconomic rights and privileges and those who fail are condemned to a continuous process of normalization, left vulnerable to arbitrary domestic power.

In short, assessment as a normalizing process makes clear one of the biggest incongruities in AL literature: if the candidate performs a task proficiently and achieves its objectives, why should institutions punish their performance due to the presence of other spoken languages in their performances? The very technical concept of proficiency, as explained above, brings to light assessment as a ritual for punishing identities and, consequently, maintaining social privileges.

Taking into account Foucault's concepts of norms and normalization, the next section presents the methodological path that guided the analysis of the criteria for assessing phonological control in the CEFR.

5. Methodological procedures

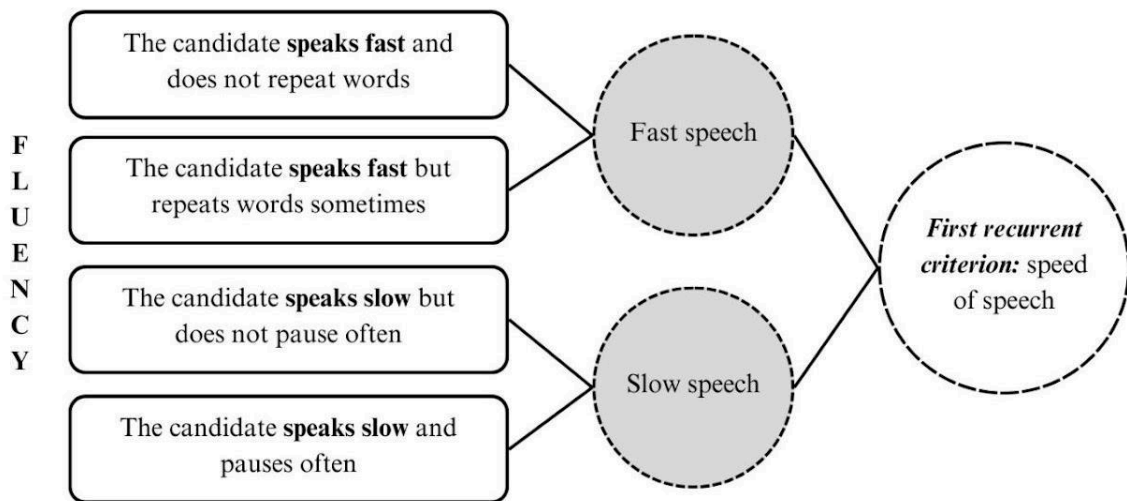
In order to analyze the assessment criteria present in the CEFR scale for phonological control, this paper presents an original adaptation of Fulcher, Davidson & Kemp's (2011) methodology for selecting assessment criteria and developing rating scales referred as PDTs or Empirically derived, Binary Choice and Boundary Definition scales (EBBs). According to Fulcher, Davidson & Kemp (2011: 9), their methodology

[...] is set forth as a series of repeated and branching binary decisions. EBBs are constructed by rank ordering performances on test tasks and then identifying key features that judges use to separate the performances into adjacent levels. EBBs represent an innovation in the logic of how raters judge performance with reference to performance data in specific contexts of language use. EBBs [...] are relatively easy to use in real-time rating, and do not place a heavy burden on the memory of the raters.

As can be seen, Fulcher, Davidson & Kemp (2011) define their methodology as qualitative-interpretative, empirically based, since exam developers consider oral performance data and the expertise of assessors and/or experts for the selection and ranking of assessment criteria. Although PDTs are empirical methodologies for developing criteria and constructing rating scales, they can also be useful in deconstructing the same scales by isolating the constitutive elements of the rating criteria present in their descriptors.

In this context, we began the analysis process with the proficiency scale of interest. The first procedure is to look at the recurring criteria between descriptors in order to identify the criteria that limit the classifications of satisfactory and unsatisfactory candidate performances. Figure 2 below illustrates the first analytical step, which we have adopted in this study, exemplified by the analysis of a hypothetical scale for assessing fluency.

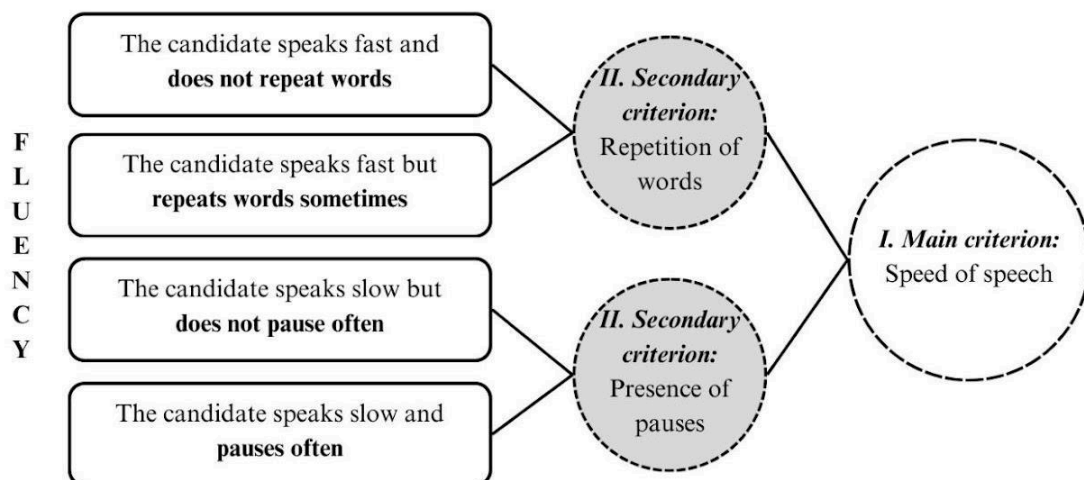
Figure 2 - Step 1 for isolating criteria.



Source: Elaborated by the author.

As illustrated in Figure 2, considering a four-level proficiency scale for assessing fluency, we isolate the criterion that limits satisfactory and unsatisfactory performances - the speed of speech. Likewise, we continued the analysis by isolating the criteria that limit the two highest and lowest ranges of the scale, in order to identify the criterion that determines the limits between them. Figure 3 below illustrates this process.

Figure 3 - Step 2 for isolating criteria.



Source: Elaborated by the author.

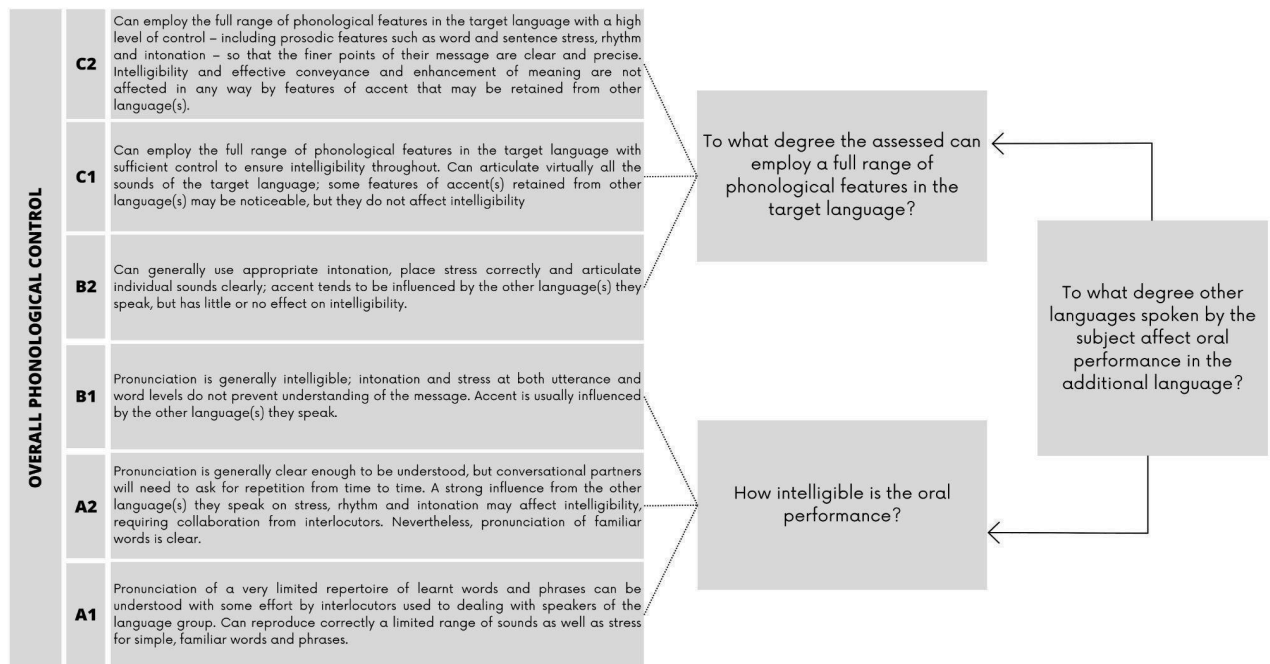
As illustrated above, by analyzing the recurrence and hierarchization of criteria, it is possible to reconstruct the operationalization of the fluency construct on this hypothetical scale. In this instrument, the speed of speech is what determines the success or failure of the candidate in the assessment process, while the repetition of words limits the more proficient levels and pauses limit the less proficient performances. The number of branches of the PDTs thus depends on the number of levels presented in the proficiency scale to be analyzed and researchers' interpretation and expertise analyzing descriptors are fundamental for establishing hierarchical relationships between criteria.

Given the methodological procedures employed in this paper in order to analyze rating scales, the next section presents results of the application of PDTs in the CEFR's phonological control assessment scale, as well as fosters discussions about the normalizing potential of its operational definition.

6. Analysis and discussion

The CEFR's phonological control assessment scale comprehends six bands, which describe levels comprising A1 as the least proficient performance and C2 as the most proficient performance. As described in the methodology section of this paper, initially, recurrent criteria that divide performances according to their proficiency level were isolated. Then, we conducted the same process to identify the criteria that limit levels A1, A2, B1 and B2, C1 and C2, as Figure 4 illustrates below.

Figure 4 - Criteria for overall phonological control from CEFR.



Source: Elaborated by the author.

As illustrated in Figure 4, the main criterion for rating oral performances is the intrusion of “language features” retained from other languages into the pronunciation of the target language. Next, at the three highest levels of the scale, segmental and suprasegmental aspects of the target language's phonological system is what distinguishes higher proficiency levels. Regarding the three lowest levels, intelligibility of oral performances sets limits that differentiate the three most basic ranges. With regard to the criteria extracted from the scale, we observed that both the CEFR and the specialized literature validate the PDT with regard to the operational definition of phonological control and the hierarchization of its assessment criteria. Interestingly, the Council of Europe (2020) provides its own definition of intelligibility, centering it on the figure of the interlocutor who, in this case, is represented by the assessor: “intelligibility, how much effort is required from the interlocutor to decode the speaker’s message; the extent of influence from other languages spoken; control of sounds [...]” (Council of Europe 2020: 133).

Additionally, intelligibility is associated with the production of segmental elements - that is, the assessor's perception of what characterizes an accent that is easy or difficult to understand, what is referred to in the literature as the listener-centered intelligibility issue

(Fulcher 2010; Thomson 2018; among others). The role of phonetic segments in performance classification is distant from the progressive proposal of the Council of Europe (2020) itself on language and assessment policies.

In this way, penalizing variation in the pronunciation of segments based on the perception of intelligibility centered on the assessor, without providing a clear definition of standards to be achieved or reproduced by the candidates, creates conditions for the potential promotion of social injustices. Based on this, the presence of elements retained from other languages in oral performances can arbitrarily determine candidates' access to job and education opportunities, as well as international mobility, which for us is a clear parallel between Foucault's (1987, 2007) concept of norms as unattainable imposed ideals for social control.

In view of this scenario, table 1 below presents important issues we raise in relation to the hierarchization of the CEFR's assessment criteria for phonological control.

Table 1 - Operationalization issues in the CEFR scale for overall phonological control.

Band	Excerpts of descriptors	Operationalization issue
C2	Intelligibility and effective conveyance and enhancement of meaning are not affected in any way by features of accent that may be retained from other language(s).	Segmental and suprasegmental retained from other languages are not apparent to the assessor in oral performances. The absence of elements that evidences the candidate's foreign origin guarantees their higher ranking, which depicts a view of language as a homogeneous phonological structure.
C1, B2	Can articulate virtually all the sounds of the target language; some features of accent(s) retained from other language(s) may be noticeable, but they do not affect intelligibility. Accent tends to be influenced by the other language(s) they speak, but has little or no effect on intelligibility.	Although intelligibility is not affected, the subjective perception of an "accent" possibly retained from other languages downgrades the candidate's performance to a lower satisfactory position. The "intrusion" of other languages as a descriptor of the second lowest level of the scale illustrates that candidates classified as B2 demonstrate proficiency performing a given language task, but their foreign background is still noticeable.
B1	Accent is usually influenced by the other language(s) they speak.	Unlike the previous cases, intrusion of elements retained from other languages spoken by candidates is evident. This aspect thus defines a boundary between satisfactory and unsatisfactory performances. Norms, in this case, marginalize the acceptable performance to the lower level of proficiency bands.
A2, A1	A strong influence from the other language(s) they speak on stress, rhythm and intonation may affect intelligibility, requiring collaboration from interlocutors. Nevertheless, pronunciation of familiar words is clear. Can reproduce correctly a limited range of sounds as well as stress for simple, familiar words and phrases.	In the lower bands of the scale, candidates' performance has a clear influence of other languages and it becomes evident to the assessor (who can even "correct" the assessed as a form of "collaboration"). Candidates' ability to reproduce sounds considered to belong to ideal norms still distinguish lower proficiency bands, which evidences that the greater the presence of foreignness in linguistic performances, the greater the potential penalty suffered by candidates.

Source: Elaborated by the author.

As can be observed, despite Council of Europe propaganda (2020), what is observable in the CEFR descriptors is a very distinct reality: the "intrusion" of elements retained from other languages, even if they do not affect the candidate's proficiency in performing language tasks, is enough reason for penalties. In this respect, what is evident is the operational definition of language proficiency based on formalist, homogenous, purist and ideal models (Oliveira 2014, Camargo 2016). Language assessment, in this sense, can potentially assume normalizing traits and, in broader political scenarios, can hinder the promotion of fair conditions for social, economic and educational opportunities.

In conclusion, the PDTs and the hierarchization of segmental and suprasegmental aspects in the description of the different levels of proficiency demonstrates that the CEFR and the Council of Europe's (2020) language policies can be employed as normalizing dynamics, possibly determining the arbitrary penalization of linguistic varieties and their speakers. In addition, descriptors that define the most and least proficient performances can serve as arguments justifying the promotion of social injustices and restrict foreigners' access to basic human rights, such as international mobility.

7. Conclusions

In this paper, we examined assessment criteria contained in the CEFR, developed by the Council of Europe (2020), for assessing phonological control in oral performances in foreign languages. We observed the presence of two main linguistic aspects for classifying candidates: the production of segmental elements, like vowels and consonants, as well as the employment of suprasegmental elements, like word stress, intonation, tone, and rhythm. From the development of the PDTs, we noticed that the CEFR hierarchically organizes segmental criteria related to the presence of linguistic aspects retained from other languages spoken by the candidate in such a way as to progressively punish the speaker's foreign accent.

The progressive punishment of the presence of segmental aspects related to other languages spoken by the candidate highlights the potential normalizing dynamic of assessment processes, since it promotes an ideal of linguistic homogeneity and arbitrarily punishes varieties that do not belong to ideal norms. Furthermore, the establishment of ideals of nativeness for the classification of candidates enables assessors and educational institutions to promote social injustices, thus hindering foreigners' access to international mobility.

We hope to stimulate the interest of applied linguists who investigate language assessment about its social impact, since assessment results guide decision-making that not only affects candidates' daily lives, but also their access to job opportunities, education and, in some cases, even survival. In addition, we expect the discussions about assessment as a normalization process to set precedents for the development of studies that are concerned with other assessment instruments and policies with global impact.

Finally, this study is limited in its scope, since we dedicated our analysis to identify potential operational issues in the phonological control rating scale developed by the Council

of Europe (2020), thus not reflecting the breadth and complexity of the CEFR operationalization of language proficiency entirely.

References

- Bundgaard-Nielsen, R. L.; Best, C. T.; Kroos, C.; Tyler, M. D. 2012. Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility. *Applied Psycholinguistics* 33(3): 643-664.
- Çagatay, S.; Gürocak, F. U. 2016. Is CEFR Really Over There? *Procedia - Social and Behavioral Sciences* 232: 705-712.
- Camargo, H. R. E. 2016. Uma Língua Inglesa para chamar de minha: equívocos sobre o bom falante de inglês. *Revista Estudos Linguísticos* 45(2): 651-665.
- Council of Europe. 2020. *Common European Framework of Reference for Languages: Learning, teaching and assessment*. Strasbourg, France: Council of Europe Publishing.
- Darvin, R.; Norton, B. 2015. *Identity and a model of investment in Applied Linguistics*. Cambridge: Cambridge University Press.
- Darvin, R.; Norton, B. 2017. Language, Identity, and Investment in the Twenty-first Century. In *Language Policy and Political Issues in Education* (227-240). *Encyclopedia of Language and Education*. Springer International Publishing.
- Davies, A. 2008. *Assessing Academic English: Testing English proficiency 1950–89: The IELTS solution*. Cambridge: Cambridge University Press.
- Foucault, M. 1987. *Vigiar e Punir: história da violência nas prisões*. Petrópolis: Editora Vozes.
- Foucault, M. 1995. *Discipline and Punish: The Birth of the Prison*. Second edition. New York: Vintage Books.
- Foucault, M. 2007. *Security, Territory, Population*. Basingstoke: Palgrave Macmillan.
- Fulcher, G. 2010. *Practical Language Testing*. London: Hodder Education/Routledge.
- Fulcher, G. 2015. *Re-examining Language Testing: A Philosophical and Social Inquiry*. London & New York: Routledge.
- Fulcher, G.; Davidson, F. 2007. *Language Testing and Assessment: An Advanced Resource Book*. London & New York: Routledge.
- Fulcher, G.; Davidson, F.; Kemp, J. 2011. Effective rating scale development for speaking tests: Performance decision trees. *Language Testing* 28(1): 5-29.
- Fulcher, G.; Harding, L. (Eds.). 2021. *The Routledge Handbook of Language Testing*. London: Routledge. <https://doi.org/10.4324/9781003220756>.
- Gooch, R.; Saito, K.; Lyster, R. 2016. Effects of recasts and prompts on L2 pronunciation development: Teaching English /r/ to Korean adult learners. *System Journal* 60(1): 117-127.
- Harding, L.; Mcnamara, T. F. 2018. Language assessment: The challenge of ELF. In *Routledge Handbook of English as a Lingua Franca* (1st ed., 570-582). Abingdon & New York: Routledge.
- Julkowska, I. A.; Cebrian, J. 2015. Effects of listener factors and stimulus properties on the intelligibility, comprehensibility and accentedness of L2 speech. *Journal of Second Language Pronunciation* 1(2): 211-237.
- Kelly, M. 2019. What's In a Norm? Foucault's Conceptualisation and Genealogy of the Norm. *Foucault Studies* 27: 1-22. <https://doi.org/10.22439/fs.v27i27.5889>.

- Kennedy, S.; Trofimovich, P. 2008. Intelligibility, comprehensibility and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review* 64(3): 459-489.
- Machado, F. F.; Delfino, M. S. L.; Rodrigues, C. O. 2021. Olhares decoloniais para o ensino de inglês. *Revista Primeira Escrita* 8(1): 115-129.
- Mcnamara, T. 2011. Managing learning: Authority and language assessment. *Language Teaching* 44(1): 500-515.
- Mohammed, N. A.; Raof, A. H. A.; Yusof, M. A. 2021. Challenges in Implementing the CEFR in English Language Classrooms. *LSP International Journal* 8(2): 115-128. <https://doi.org/10.11113/lspi.v8.17977>.
- O'Brien, M. G. 2021. Ease and Difficulty in L2 Pronunciation Teaching: A Mini-Review. *Front. Commun* 5: 626985. <https://doi.org/10.3389/fcomm.2020.626985>.
- Oliveira, L. A. 2014. Métodos de ensino de inglês: teorias, práticas, ideologias. São Paulo: Parábola Editorial.
- Pennycook, A. 2017. *The cultural politics of English as an international language*. London and New York: Routledge.
- Segat, G. L.; Sarmiento, S. 2022. Avaliação linguística em periódicos de Linguística Aplicada no Brasil: mapeamento das publicações. *Revista Horizontes de Linguística Aplicada* 21(2): 1-20.
- Spolsky, B. 1995. *Measured words*. Oxford: Oxford University Press.
- Thomson, R. 2018. Measurement of accentedness, intelligibility, and comprehensibility. In *Assessment in Second Language Pronunciation* (1st ed., 11-29). New York: Routledge.
- Zaki, A. W.; Darmi, R. 2021. The implementation of CEFR in ESL Learning: Why does it matter to the Malaysian Education System? *Asian Journal of Assessment in Teaching and Learning* 11(2): 1-13.